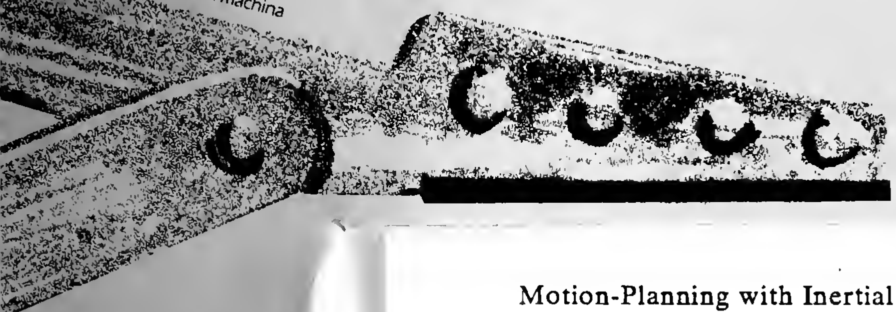
Robotics Research Technical Report

[G]eneratorium omnis laboris ex machina



Motion-Planning with Inertial Constraints

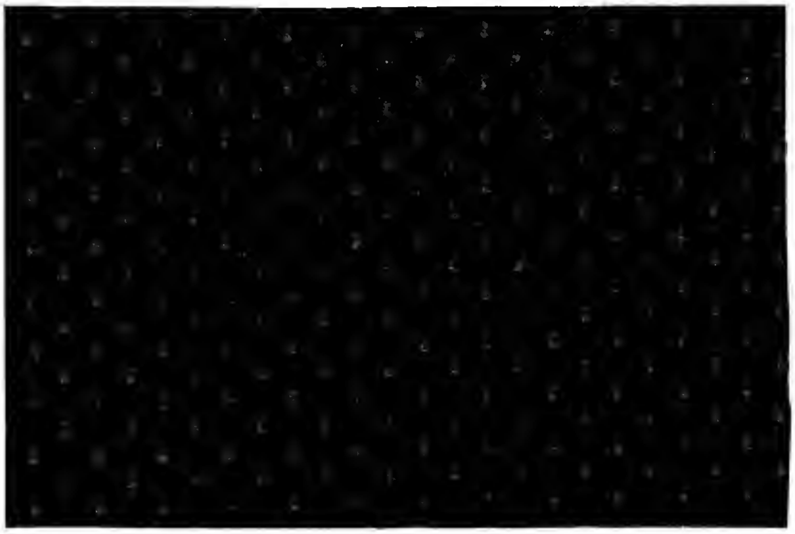
by

Colm Ó'Dúnlaing

Technical Report No.230
Robotics Report No. 73
July, 1986

New York University
[Couran]t Institute of Mathematical Sciences

Computer Science Division
251 Mercer Street New York, N.Y. 10012



Motion-Planning with Inertial Constraints

by

Colm Ó'Dúnlaing

Technical Report No.230
Robotics Report No. 73
July, 1986

New York University
Dept. of Computer Science
Courant Institute of Mathematical Sciences
251 Mercer Street
New York, New York 10012

This work was supported in part by the National Science Foundation under grant DCR 84-01898.

Motion-planning with inertial constraints[1]

Colm Ó'Dúnlaing
Courant Institute of Mathematical Sciences
251 Mercer Street
New York, New York 10012

A body B must move from a placement Z_0 to a placement Z_1 while avoiding collision with a set S of moving obstacles. The motion must satisfy an inertial constraint: the acceleration cannot exceed a given bound M. The problem is analyzed, and polynomial-time motion-planning algorithms are given, for the case of a particle moving in one dimension.

[1]This work was supported in part by the National Science Foundation under grant DCR 84-01898.

1. Introduction.

There has been much recent research into many motion-planning problems as computational problems, but there has been little work done on the problem of motion-planning in the presence of constraints on velocity or acceleration.

The relevance of such studies is rendered plausible in the following scenario: an automated manufacturing system is in the process of design, and all the steps in the manufacture have been planned as a sequence of motions of various robots, feeder belts, and so forth, but the overall speed of assembly has yet to be planned. It is likely in certain instances that this overall speed be limited by the fragility of the parts under assembly; the speed of assembly, and consequently the throughput of the line, should be chosen to be as rapid as possible subject to limits on the inertial forces brought to bear on the parts being assembled. Therefore this paper addresses the motion-planning problem with inertial constraints, in the following simple case:

> To decide whether it is possible for the particle to follow a (space-time) path with boundary positions and velocities (x_a, v_a) and (x_b, v_b) at boundary times a and b, respectively, where the particle evades two 'pursuing' particles, whose positions are given by functions $f(t)$ and $g(t)$ respectively, where f and g are parabolic splines, and the acceleration of the particle in question never exceeds a fixed bound M in absolute value.

It is curious that, even for the simple case studied here, a satisfactory solution to the constrained problem remains nontrivial. This paper provides two alternative solutions to the above problem: one with a runtime of $O(n^2)$ and a variant which has a better runtime of $O(n\log(n))$, where n is the total complexity of the pursuit functions, measured by the total number of times they change acceleration.

Here is an outline of the strategy followed in this paper. First, the problem can be stated in a more useful form: given initial position and velocity $p = (x, v)$, to compute the set $R(p)$ of final positions and velocities which can be reached from the given initial conditions by a motion which satisfies the given constraints.

It will be useful therefore to consider many transformations from points in *phase space* to sets of points in phase space, such as the mapping R just introduced. Here 'phase space' denotes the set of all ordered pairs (x, v) of real numbers where x is position and v is velocity. Our goal will be to compute a value $R(p)$ rather than the functional relation $R()$. (The latter problem seems intrinsically more complex.) It will, however, be very useful to compute unions of sets $R(p)$: accordingly, the problem is restated in the following slightly more general form.

Given a *set* I of initial conditions, to determine the set $R(I)$ of final points reachable from this set by paths which obey the given constraints.

It is to be emphasized that the problem as formulated is existential; given I, to compute the set of points q for which *there exists* a point p in I such that q is in $R(p)$. Our method will apply to a certain family of convex sets I in phase space, which we shall call (for want of a better term) *standard phase-regions*.

The paper is organized as follows: first, we study the problem of finding a minimum-acceleration path, given initial and final positions and velocities, in the absence of any other constraints. This path turns out to be a spline function (of position against time) in which the acceleration is suddenly reversed at one point but otherwise remains constant: hence we shall call the solution the 'bang-bang interpolant.' Next we shall study the feasible region $FR(p)$ accessible from a given point p in the absence of pursuer functions but with bounds of $\pm M$ on the acceleration, and discover that the region is convex and bounded by two parabolic segments in phase space. Next, we consider the problem of computing the set of points in phase space reachable by *some* trajectory which avoids two pursuer functions f and g which are assumed to be a 'reachable pair' (a term defined in the paper), where there are no limitations on the initial conditions. Next, we study how to compute what we call the 'reachable margins' determined by two pursuit functions: this enables us to solve the original problem as a special case. Finally, we consider the more general problem of finding the set $R(I)$ of points reachable from a *set* of points I; with suitable restrictions on I we obtain an alternative method of solving the original problem, in overall time $O(n\log(n))$. The method here is to compute iteratively the sets of points reachable at various time-points along the interval, such that between the time points the pursuit functions have fixed acceleration.[2] A concluding section discusses the limitations imposed on the problem here and possibilities for generalization.

2. Definitions and terms.

We consider a particle moving on the real line, whose *trajectory* $x(t)$ is subject to some constraints. The set of pairs (t,x), where t is time and x is position, will be called (somewhat grandly) *space-time*. The set of triples (t,x,v) where in addition v denotes velocity will be called *parametrized phase-space*.

A *trajectory* $x(t)$ is a curve with bounded, piecewise continuous second derivative. Its domain of definition will be some closed time-interval $[a,b]$ where $a < b$. A *quadratic spline* is a trajectory with continuous derivative and piecewise constant second derivative.

[2]The author is indebted to Alan Siegel for proposing and discussing this approach.

Suppose that we are given two *pursuit functions* f and g defined on $[a,b]$ where $f(t) \le g(t)$ everywhere and both these functions are quadratic splines. Suppose that a positive real number M is also supplied: it will serve as an *inertial bound*.

Given these 'pursuit functions,' the inertial bound M, initial position and velocity (x_a, v_a) and final position and velocity (x_b, v_b), the problem is to determine whether there exists a trajectory $x(t)$ satisfying the following relations:

(boundary conditions)

$$x(t) = x_t;\ \dot{x}(t) = v_t, \quad t = a, b; \tag{2.0}$$

(pursuit constraints)

$$f(t) \le x(t) \le g(t), \quad t \in [a,b]; \tag{2.1}$$

and (inertial constraint)

$$|\ddot{x}(t)| \le M, \quad t \in [a,b]. \tag{2.2}$$

A trajectory satisfying the inertial constraint (2.2) will sometimes be called an *accleration-bounded* trajectory for short. A trajectory satisfying the latter two constraints (2.1), (2.2) will be called an *admissible* trajectory. An admissible trajectory also satisfying the boundary conditions (2.0) will be called an *admissible solution* to (2.0).

Let a be a fixed point in time. A *phase region* at a will be a set of points (a, x, v) in parametrized phase-space. Corresponding to such a region I, we write

$$R_{a,b}(I)$$

to denote the set of points (b, x_b, v_b) at time b for which there exists an admissible trajectory $x(t)$ beginning at some point in I and ending at (b, x_b, v_b). We call this the set of points *reachable* from I. Again, given such a region I, we write

$$FR_{a,b}(I)$$

to denote the set of points (b, x_b, v_b) 'freely reachable from I' − reachable by a trajectory $x(t)$ which begins at some point in I and satisfies the inertial constraint (2.2) but which ignores the constraint (2.1).

Sometimes, by abuse of notation, we shall consider a phase region to be just a subset of phase space, i.e., a set of ordered pairs (x, v) rather than ordered triples (a, x, v).

A *uniform-acceleration* trajectory is one of the form

$$x(t) = G + \tfrac{1}{2}N(t-H)^2. \tag{2.3}$$

Here the acceleration N is fixed. Given initial conditions (a, x_a, v_a), there is exactly one uniform-acceleration trajectory with acceleration N passing through this point, and we denote it as $\Phi(a, x_a, v_a)$ (its dependence on the parameter N is left implicit). It is convenient to parametrize a uniform-acceleration trajectory of the form (2.3) by the coordinates (H, G) in space-time of its apex. Under this correspondence it is straightforward to verify that

$$\Phi^{-1}(H,G) = (a, G + \tfrac{1}{2}N(a-H)^2, N(a-H)),$$

and it is easy to invert the above relation and show that

$$\Phi(a, x_a, v_a) = (H, G) = \left(a - \frac{v_a}{N}, x_a - \frac{v_a^{\,2}}{2N}\right). \tag{2.4}$$

One should note that the constraints (2.1) are only assumed to apply over their domain of definition, to make the problem as general as possible. Thus one can consider the pursuit functions to have discontinuous and unbounded jumps at the endpoints a and b. To capture this notion, we shall use F (respectively, G) to denote the set of all points (t,x) bounded vertically above (respectively, below) by the graph of f (respectively, the graph of g). Thus,

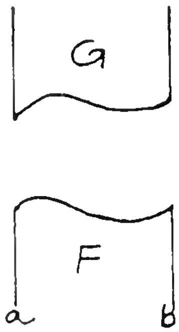
$$F = \{(t,x) : a \le t \le b \text{ and } x \le f(t)\},$$

and

$$G = \{(t,x) : a \le t \le b \text{ and } x \ge g(t)\}. \tag{2.5}$$

Thus an acceleration-bounded trajectory is admissible if and only if it does not meet the interior of F or of G.

Orientation of axes. In all diagrams of phase-space and space-time in this paper, the x-axis will be vertical and the other will be horizontal. The terms 'earlier' and 'left of' mean with respect to time; similarly, 'later' and 'right of.' Sometimes alternative positional



Illustrating the sets F and G

descriptions will be used for brevity: for instance, in phase-space (with the v-axis implicitly horizontal) (x,v) is northwest of (x',v') if and only if $x \geq x'$ and $v \leq v'$.

3. The bang-bang interpolant, phase space, and $FR(p)$

In this section we consider the set of all trajectories with prescribed positions and velocities at time-points a and b: the problem is to compute an optimal-acceleration trajectory. This problem is similar to one recently studied by R. Farwig and D. Zwick.[1] We shall then use the solution to define analytically the phase region $FR_{a,b}(p)$ at time b consisting of all points (x,v) reachable from initial position and velocity p at time a by a trajectories which respect the inertial constraint (2.2).

In the unconstrained case where f and g are effectively infinite, it is possible to compute a path $x(t)$ which respects the boundary conditions (2.0) and achieves *minimax* acceleration, i.e., one in which the quantity

$$\sup \{ |\ddot{x}(t)| : t \in [a,b] \} \tag{3.0}$$

is minimized; moreover, the solution is unique:

Lemma 1. (Principle of the 'bang-bang' interpolant) Given parameters (x_a, v_a, x_b, v_b), there exists a minimax acceleration solution $x(t)$ consisting of two parabolas spliced together, for which the (constant) second derivatives in each segment are equal in magnitude but opposite in sign. The minimax solution in this case is unique.

Proof. First we consider a solution curve whose *velocity diagram* is formed of two linear segments, and thus has an instantaneous change in acceleration at time τ, say, but which is otherwise continuous. The distance and velocity constraints require that this curve should pass through (a,v_a) and (b,v_b), and that the point (τ,v_τ) where the acceleration changes should be on a fixed line parallel to the line joining the endpoints (this is to satisfy the integration constraints (2.0)). It is intuitively clear that among all such curves the minimum-acceleration curve is the one for which these two accelerations are equal in magnitude but opposite in sign: this can be justified formally by a simple but tedious calculation. See the accompanying figure. Let us call the curve defined in this way the 'bang-bang interpolant.'

Now let $w(t)$ be any other solution, and suppose that it achieves a better minimax acceleration. Without loss of generality, the velocity v_τ is greater than $\max(v_a,v_b)$ (otherwise, modify the argument by considering the related optimization problem where the constraints are reversed in sign). Let $\alpha(t)$ represent $\ddot{w}(t)$; by assumption, $\alpha(t)$ is defined and continuous almost everywhere. Now one can argue that it is impossible that $|\alpha(t)| < |\dot{v}_t|$

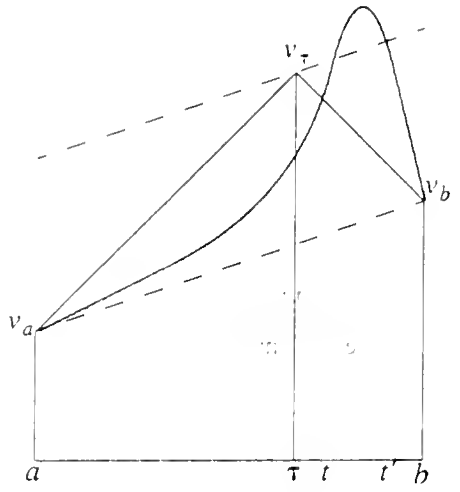


Figure: Illustrating Lemma 1.

(almost everywhere): for suppose that $\alpha(t) < \dot{v}_t$ for almost all $t \leq \tau$; it follows that $w(t) < v_t$ for all nonzero $t \leq \tau$. However, v and w both integrate to the same value over $[a,b]$. Hence, for some $t > \tau$, $v_t < w(t)$. Were $\alpha(t') > \dot{v}_t$ for all $t' > t$, $w(b)$ would exceed v_b, violating the boundary conditions. Hence for some $t' > \tau$, $\alpha(t') < \dot{v}(t') < 0$. More significantly, the set of such points t' has nonzero measure. Q.E.D.

The following lemma is proved in an appendix to this paper.

Lemma 2. The quantity v_τ, which serves to define the bang-bang interpolant of Lemma 1, satisfies the equation

$$(v_\tau - C)^2 = \frac{(v_a - C)^2 + (v_b - C)^2}{2}. \qquad (3.1))$$

where

$$\dot{C} = \frac{x_b - x_a}{b-a}. \qquad (3.2)$$

Furthermore μ, the initial acceleration, satisfies

$$\mu = \frac{2v_\tau - v_a - v_b}{b-a}. \qquad (3.3)$$

The quantity τ introduced in Lemma 1 is deducible from v_τ, and except in the case where $v_a + v_b = 2C$, exactly one of the two solutions to (3.1) implies a legal value for τ, i.e., τ in $[a,b]$; the 'degenerate' case has a unique solution which is parabolic (i.e., the acceleration is constant). Specifically,

$$v_\tau = C - s\sqrt{\tfrac{1}{2}((v_a-C)^2+(v_b-C)^2)}, \tag{3.4}$$

where the parameter $s = +1$ if $v_a - C \neq C - v_b$ and the maximum of the two quantities $|v_a - C|$, $|v_b - C|$ is achieved by one which is positive: i.e.,

$$\max(v_a - C, v_b - C) = \max(|v_a - C|, |v_b - C|);$$

otherwise, if the two quantities are not equal in absolute value but opposite in sign, then $s = -1$. (If the quantities are opposite then the case is degenerate, the quantities τ and v_τ are irrelevant, and there is a unique optimal solution in which the acceleration is constant throughout.) •

Now we want to exploit Lemma 2 to solve the following problem: given initial time, position, and velocity a, x_a, and v_a, respectively, a fixed time $t > a$, and a bound M on the absolute value of the acceleration, to describe the set $FR_{a,t}(x_a, v_a)$, namely, the region of points (x, v) in 'phase space' such that there exists a trajectory through (a, x_a, v_a) and (t, x, v) satisfying the inertial bound

$$|\ddot{x}(s)| \leq M, s \in [a, t].$$

To begin with, and with little loss of generality, we shall make the simplifying assumption that a, x_a and v_a are all zero, and hence t is positive. Clearly, if there exists a solution satisfying the given inertial constraints, then the bang-bang interpolant satisfies the constraints. Therefore the above inequality translates into

$$-Mt^2 \leq 2x - vt - s\sqrt{2x^2 + 2(vt-x)^2} \leq Mt^2, \tag{3.5}$$

where s is ± 1; let us simplify the conditions determining the sign of s in this case. According to Lemma 2, s is positive if the one of the two quantities

$$v - x/t, \ -x/t$$

which has larger absolute value is positive. Since t is positive we can apply the same criterion to the quantities

$$vt - x \text{ and } -x.$$

For s to be positive, if $vt - x$ is negative then we require that x be negative and have larger absolute value; equivalently, $x \leq vt - x$ or $2x - vt \leq 0$; whereas, if $-x$ is negative (so x is positive), we require that $vt - x$ be positive and have larger absolute value; i.e., $x \leq vt - x$ again or $2x - vt \leq 0$. The choice of s is ambiguous when the quantities are equal in magnitude but opposite in sign, which happens when $2x - vt = 0$. Thus s is $+1$ if $2x - vt$ is negative, -1 if $2x - vt$ is positive, and indeterminate if $2x - vt$ is zero. The inequality (3.5)

defines the set $FR_{0,t}(0,0)$ by an implicit relation between x and v parametrized by M and t. Let us split (3.5) into the two cases:

Case (i)

$s=+1$, so $2x-vt \le 0$. Then since the term between the two inequalities is nonpositive, the second inequality is automatically satisfied and the important one is the first one, which may be rewritten in the following form:

$$2x-vt+Mt^2 \ge \sqrt{(2x-vt)^2+v^2t^2}.$$

Since the right-hand side is positive, this inequality may be preserved when squaring both sides and we obtain

$$(2x-vt)^2+4Mt^2x-2Mvt^3+M^2t^4 \ge (2x-vt)^2+v^2t^2,$$

so, cancelling the term $(2x-vt)^2$, bringing the terms involving v to the right-hand side, cancelling throughout a common factor of t^2, and adding Mt^2 to complete the square, we obtain

$$4Mx+2M^2t^2 \ge (v+Mt)^2$$

or

$$4Mx \ge (v+Mt)^2-2M^2t^2. \tag{3.6}$$

Case (ii)

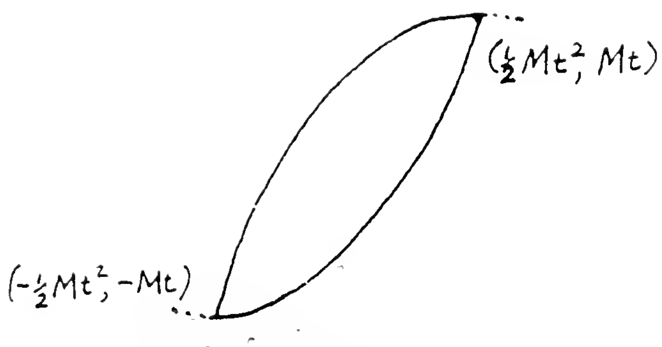
$s=-1$, so $2x-vt \ge 0$. Reasoning in the same way, the left-hand inequality becomes redundant and the right-hand inequality becomes

$$4Mx \le 2M^2t^2-(v-Mt)^2. \tag{3.7}$$

Summarizing, the region $FR_{0,t}(0,0)$ may be expressed as the union of two regions, each region being bounded by a parabolic segment and a straight-line segment (the straight line satisfying the equation $2x-vt = 0$). A straightforward calculation reveals that both parabolas intersect the straight line $2x = vt$ at the same two points in phase space, namely (as is hardly surprising, since they represent the points reached by maintaining acceleration $\pm M$) $\pm(\frac{1}{2}Mt^2, Mt)$. Therefore the region $FR_{0,t}(0,0)$ may be expressed more succinctly as the region bounded by two parabolic segments in the (x,v)-plane:

$$FR_{0,t}(0,0) = \{(x,v):(v+Mt)^2-2M^2t^2 \le 4Mx \le 2M^2t^2-(v-Mt)^2\}. \tag{3.8}$$

It is trivial to show that $FR_{a,t}(0,0) = FR_{0,t-a}(0,0)$, which trivially generalizes equation (3.8) to nonzero values of a. If $x(s)$ is an acceleration-bounded trajectory (i.e., one



Illustrating $FR_{0,t}(0,0)$

satisfying the constraint (2.2)) passing through (a, x_a, v_a), then $x(s)-x_a-(s-a)v_a$ is an acceleration-bounded trajectory passing through $(a, 0, 0)$. This, combined with the obvious inverse transformation, establishes a bijection between the set of acceleration-bounded trajectories through (a, x_a, v_a) and $(a, 0, 0)$ respectively, and therefore we conclude

$$FR_{a,t}(x_a, v_a) = \{ (x_a+(t-a)v_a, v_a)+(x,v) : (x,v) \in FR_{0,t-a}(0,0) \}. \tag{3.9}$$

We can express this more succinctly using the notation of Minkowski sum of sets of vectors (given two sets X and Y of vectors, the Minkowski sum $X+Y$ is defined as $\{x+y : x \in X, y \in Y\}$). Let $T_{a,t}$ be the linear transformation which takes a point (x_a, v_a) to $(x_a+(t-a)v_a, v_a)$. Then, for any phase region I at time a,

$$FR_{a,t}(I) = T_{a,t}(I)+FR_{0,t-a}(0,0). \tag{3.10}$$

4. Phase region formed from a reachable pair of pursuit functions.

The *phase region at time a* formed from a pair (f,g) of pursuit functions is the set of all points (a,x,v) in phase space at time a such that there exists an admissible trajectory beginning at (a,x,v). Similarly the phase region at time b is the set of all endpoints of admissible trajectories.

In this section we shall construct explicitly the phase region (at time b) formed from a pair of pursuit functions assuming they are 'reachable.'

Definition. A point (τ, x_τ, v_τ) in parametrized phase space at time τ (where $a<\tau\leq b$) is *reachable* if there exists a trajectory $x(t)$ defined over the interval $[a,\tau]$, satisfying the pursuit and inertial constraints (2.1) and (2.2), and ending at the point (τ, x_τ, v_τ). A pair (f,g) of pursuit functions is *reachable up to time c* if $a<c\leq b$ and for every τ in $(a,c]$, both $(\tau, f(\tau), \dot{f}(\tau))$ and $(\tau, g(\tau), \dot{g}(\tau))$ are reachable. (Usually c will be left implicit.)

Given a point $p = (\tau, x, v)$, there are unique trajectories υ_p and δ_p, with acceleration M and $-M$ respectively, passing through p. (The notation has been chosen to connote 'up' and 'down' respectively.) Explicitly,

$$\delta_p(t) = x + v^2/2M - \tfrac{1}{2}M(t-\tau-v/M)^2,$$

and

$$\upsilon_p(t) = x - v^2/2M + \tfrac{1}{2}M(t-\tau+v/M)^2. \tag{4.1}$$

Lemma 3. (i) The trajectories υ_p and δ_p are respectively the upper and lower envelopes of all acceleration-bounded trajectories (i.e., all trajectories satisfying the constraint (2.2)) passing through p. (ii) Let τ, x, x', v, and v' be fixed, where $x \le x'$ and $v \le v'$ but they are not both equal. Let p, q, r, s denote (τ,x,v), (τ,x,v'), $(\tau,x',v$, and (τ,x',v') respectively. Then for all points $\phi<\tau$ and $\psi>\tau$, $\upsilon_p(\psi) < \upsilon_s(\psi)$, $\delta_p(\psi) < \delta_s(\psi)$, $\upsilon_q(\phi) < \upsilon_r(\phi)$, and $\delta_q(\phi) < \delta_r(\phi)$.

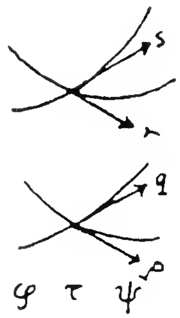
Proof. (i) Trivial. (ii) This is immediate when $x<x'$ and $v=v'$, so we need only consider the case where $x=x'$ and $v<v'$, and that follows easily from consideration of the slopes of the tangents to the four trajectories at (τ,x). •

Consider the set of points (b,x,v) where $f(b) \le x \le g(b)$. For each such point x we want to compute the set of values v such that (b,x,v) is reachable. To begin with, we compute upper and lower bounds on this set.

Definition. Given b and x as above, and F and G as in (2.5), define

$$v_u(x) = \text{Sup}\,\{v: \upsilon_{(b,x,v)} \text{ does not intersect Interior}(F)\},$$

and



Illustrating Lemma 3.

$$v_\ell(x) = \mathrm{Inf}\ \{v\colon \delta_{(b,x,v)} \text{ does not intersect Interior}(G)\}. \tag{4.2}$$

It is easy to see that these definitions are sound, i.e., that the Sup and Inf are both defined on sets of numbers which are nonempty and which have nonempty complement.

Lemma 4. If a point (b,x,v) is reachable then $v \in [v_\ell(x), v_u(x)]$.

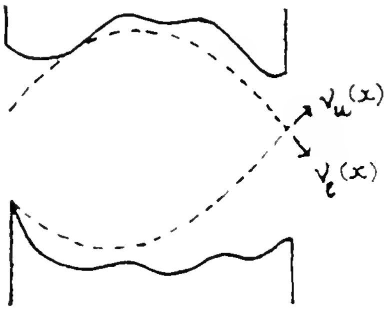
Proof. Let p denote the point (b,x,v), and suppose that p is reachable, so there is an admissible trajectory, call it γ, ending at p. Since γ intersects neither the interior of F nor the interior of G it follows from Lemma 3 that υ_p does not intersect the interior of F and δ_p does not intersect the interior of G. Thus $v \in [v_\ell(x), v_u(x)]$. •

We can give two necessary conditions that a pair f and g of pursuit functions be reachable up to time b:

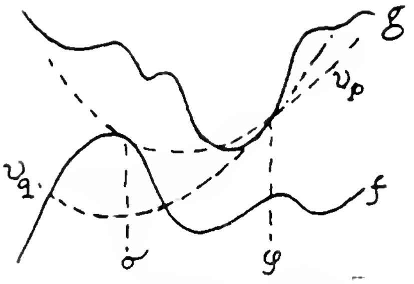
Lemma 5. Suppose that f and g are reachable up to time b. Then (i) $\ddot f \le M$ and $\ddot g \ge -M$ on $[a,b]$, and given any point $p = (\tau, x, v)$ where $\tau \in [a, b]$ and $x \in [f(\tau), g(\tau)]$, (ii) if υ_p meets F at some point σ earlier than τ then it cannot cross G (transversally) in the interval $[\sigma,\tau]$, and (iii) if δ_p meets G at some point σ earlier than τ then it cannot cross F in the interval $[\sigma,\tau]$.

Proof. (i) Suppose that $\tau \in [a,b]$. Since $(\tau, g(\tau), \dot g(\tau))$ is reachable, $\ddot g(\tau) \ge -M$ (otherwise all acceleration-bounded trajectories would penetrate G near this point); similarly $\ddot f(\tau) \le M$. (ii) Assume that such a transverse intersection exists: let ϕ be the latest point in $[\sigma,\tau]$ where the transverse intersection occurs: so $\upsilon_p(\phi) = g(\phi)$ and $\dot\upsilon_p(\phi) < \dot g(\phi)$. Let $q = (\phi, g(\phi), \dot g(\phi))$. By Lemma 3, $\upsilon_q(\sigma) < \upsilon_p(\sigma) = f(\sigma)$, and it follows, again by Lemma 3, that there exists no trajectory admissible over $[a,\phi]$ and ending at q, so q is not reachable, a contradiction; (iii) follows by parallel arguments. •

Theorem 6. Suppose that f and g are a pair of pursuit functions reachable up to time b. Then for every $x \in [f(b), g(b)]$ and for every v, (b,x,v) is reachable if and only if



Illustrating v_ℓ and v_u



Illustrating Lemma 5.

$v_\ell(x) \le v \le v_u(x)$.

Proof. Only if: Lemma 4.

If: One can show by straightforward continuity arguments that when $v = v_\ell(x)$ then $\delta_{(b,x,v)}$ meets G tangentially at some earlier point, σ, say; by Lemma 5, it does not meet f transversally between σ and b. If $\sigma = a$ then we are done. Otherwise, by the reachability hypothesis, there is a trajectory which meets g tangentially at σ and is admissible over $[a,\sigma]$; this trajectory can be spliced to the trajectory $\delta_{(b,x,v)}$ yielding an admissible trajectory ending at (b,x,v). Similarly, if $v = v_u(x)$, then (b,x,v) is reachable.

Finally, suppose that $v_\ell(x) < v < v_u(x)$. Consider the trajectory $\delta_{(b,x,v)}$. Because of the bounds on v it cannot meet the interior of G. If it never meets the interior of F then it is an admissible trajectory and we are done. Otherwise consider the family of trajectories $\upsilon_{p(\tau)}$ where $p(\tau)$ ranges over the given trajectory: $p(\tau) = (\tau, \delta_{(b,x,v)}(\tau), \dot{\delta}_{(b,x,v)}(\tau))$. We know that for $\tau = b$, $\upsilon_p(\tau)$ does not meet the interior of F, and by assumption there is an earlier time τ such that $\upsilon_p(\tau)$ meets the interior of F, so there is some intermediate point σ such that $\upsilon_p(\sigma)$ meets F tangentially. By another 'splicing' construction, invoking Lemma 5, we obtain an admissible trajectory ending at (b,x,v). **Q.E.D.**

The above theorem leads to an alternative derivation of the sets $FR_{a,b}(p)$. Suppose that $p = (a, x_a, v_a)$ is given: put $f = \dot{\delta}_p$ and $g = \upsilon_p$. Then $FR_{a,b}(p)$ is identical to the region formed from f and g at time b, they are clearly reachable, and v_ℓ and v_u defines the boundaries of the set $FR_{a,b}(p)$.

5. Constructing the reachable margins.

We shall use the results in the preceding section to solve the motion-planning problem. Our task, therefore, must be to modify the given pursuit functions f and g so that they become 'reachable' as defined in the preceding section. We shall devise an iterative method

to do this, constructing two new pursuit functions $\hat{f}$ and $\hat{g}$ with the properties that (i) they form a reachable pair up to time b, and (ii) a trajectory is admissible relative to $\hat{f}$ and $\hat{g}$ if and only if it is admissible relative to f and g. These functions f and g will be called the *reachable margins*.

This will be accomplished with the aid of two operations which we shall call 'filling' and 'shading.' Given a pursuit function f, define $\bar{f}(t)$, for $a \leq t \leq b$, by

$$\bar{f}(t) = \mathrm{Inf}\,\{x\colon \exists v(\upsilon_{(t,x,v)} \text{ does not meet Interior}(F))\};$$

similarly define

$$\bar{g}(t) = \mathrm{Sup}\,\{x\colon \exists v(\delta_{(t,x,v)} \text{ does not meet Interior}(G))\}. \tag{5.1}$$

'Filling' f or g means replacing it by $\bar{f}$ or $\bar{g}$. For this definition it is not necessary that f or g be continuously differentiable everywhere; it is enough that they be continuous everywhere and piecewise continuously differentiable.

To define the operation of shading let us begin with a simple lemma, analogous to Lemma 3.

Lemma 7. Suppose that $\ddot{g} > M$ in an interval $[c,d]$. For $c \leq \tau \leq d$ let $p(\tau) = (\tau, g(\tau), \dot{g}(\tau))$. Then for any points σ, τ, and ϕ, with $c \leq \tau < \phi \leq d$ and $\sigma \leq \tau$, $\upsilon_{p(\tau)}(\sigma) > \upsilon_{p(\phi)}(\sigma)$.

Proof. A straightforward adaptation of the method of Lemma 3: the curves $\upsilon_{p(\tau)}$ and $\upsilon_{p(\phi)}$ intersect somewhere between τ and ϕ. •

Definition. A pursuit function f or g is *simple* over an interval $[c,d] \subseteq [a,b]$ if $\ddot{f}$ (respectively, $\ddot{g}$) is constant on $[c,d]$. If g is simple over $[c,d]$, with $\ddot{g} > M$ in the interval, then by Lemma 7 there exists at most one point d' in $[c,d]$ such that $\upsilon_{(d',g(d'),\dot{g}(d'))}$ meets F tangentially at some point before d' but does not meet the interior of F in $[a,d']$. If such a

g and the 'filled' version $\bar{g}$

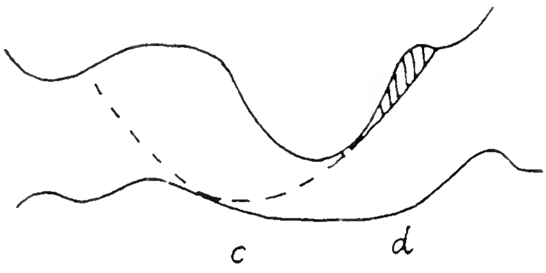
point d' exists, then the *shadow of* g against f *from the interval* $[c,d]$ is that segment of $\upsilon_{(d',g(d'),\dot{g}(d'))}$ beginning at d' and ending at the first point, if it exists, where it meets g again, or b if no such point exists. (Generally the shadow will extend beyond the endpoint d of the interval.) Similarly if f is simple in $[c,d]$ with $\ddot{f}<-M$ one can sometimes define a trajectory of the form δ_p meeting f tangentially at p in $[c,d]$ and G tangentially at some earlier point.

Notice that if the shadow, with its associated initial point d', is defined for g, then for all τ in $[c,d')$, $\upsilon_{(\tau,g(\tau),\dot{g}(\tau))}$ does not meet the interior of F to the left of τ, and for all points τ in $(d',d]$, $\upsilon_{(\tau,g(\tau),\dot{g}(\tau))}$ meets the interior of F at some point to the left of τ (Lemma 7). A symmetrically analagous property holds for the family of trajectories $\delta_{(\tau,f(\tau),\dot{f}(\tau))}$ if the shadow is defined for f in the interval.

To *shade* g relative to f over the interval $[c,d]$ means to ascertain whether the shadow exists in that interval, and, if it exists, on an interval $[d',e]$, say, to modify g by making it coincide with the shadow in that interval, otherwise leaving g unchanged. (Generally, shading g introduces exactly one point where the derivative is discontinuous.) Similarly one can shade f relative to g.

Lemma 8. (i) If $\bar{g}$ is obtained by filling g then no point above $\bar{g}$ lies in any admissible trajectory (over $[a,b]$); similarly no point below $\bar{f}$ lies in any admissible trajectory. (ii) No point above a shadow for g or below a shadow for f can be reachable.

Proof. (i) It is easy to see by definition of $\bar{g}$ that if (τ,x,v) is any point with $x>\bar{g}(\tau)$ then $\delta_{(\tau,x,v)}$ meets the interior of G and hence (Lemma 3) the point cannot be in any admissible trajectory. (ii) Suppose that (τ,x,v) is some point where (τ,x) is above a shadow for g. Let γ be any admissible trajectory ending at this point. Suppose that d' is the point at which the given shadow begins. If γ does not cross the shadow then $\gamma(d')>g(d')$ and the trajectory violates the bounds (2.1). Otherwise there must be some point σ at which it



The shadow of g against f from $[c,d]$.

crosses the shadow with a velocity greater than the velocity of the shadow at that point (because otherwise one can show it violates the inertial constraints somewhere), and hence by definition of the shadow and Lemma 3 it must cross F at some earlier point. Thus (τ, x, v) is not reachable. ●

If a shadow for g (or f) is defined beginning at some point d' then we call that part of the trajectory $\upsilon_{(d', g(d'), \dot{g}(d'))}$ (respectively, $\delta_{(d', g(d'), \dot{f}(d'))}$) ending at d' the *trail* of the shadow.

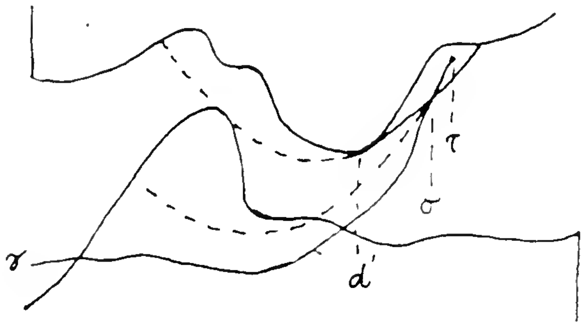
Definition. Suppose that f and g are both simple over a subinterval $[c, d]$ and both 'filled,' in the sense that $\ddot{f} \leq M$ and $\ddot{g} \geq -M$ on $[a, b]$, and they form a reachable pair up to time c. To ***extend the margins across*** d means to produce a new pair $(f^{(2)}, g^{(2)})$ of pursuit functions by the following sequence of operations:

Compute $f^{(1)}$ and $g^{(1)}$ by shading f relative to g in $[c, d]$ (respectively, shading g relative to f in $[c, d]$), and filling the resulting functions.

Compute $f^{(2)}$ and $g^{(2)}$ by shading f relative to $g^{(1)}$ and g relative to $f^{(1)}$ in $[c, d]$ and filling the resulting functions.

Theorem 9. Suppose that $\ddot{f} \leq M$ and $\ddot{g} \geq -M$ on $[a, b]$, that f and g are both simple over $[c, d] \subseteq [a, b]$, and that they form a reachable pair up to time c. Let us suppose that $f^{(2)}$ and $g^{(2)}$ are produced by extending the margins across d, as defined above. Then (i) any trajectory $x(t)$ defined over $[a, b]$ is admissible with respect to (f, g) if and only if it is admissible with respect to $(f^{(2)}, g^{(2)})$, and (ii) if $f^{(2)}$ and $g^{(2)}$ do not intersect in $[c, d]$ then they form a reachable pair up to time d.

Proof. (i) A simple consequence of Lemma 8 and the definitions. (ii) Let us first consider the case where $g^{(2)}$ and $g^{(1)}$ are different. Let γ represent the trail of the shadow for g (relative to f). In this case, γ must cross $f^{(1)}$ transversally. The function $f^{(1)}$ differed from f, therefore, by patching f with a trajectory δ with acceleration $-M$ (shading) possibly



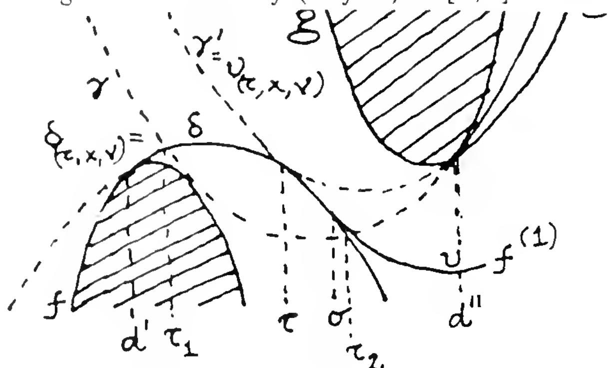
Illustrating Lemma 8

followed by a trajectory υ of acceleration M (filling). Claim that γ must actually cross δ. Otherwise it crosses υ transversally and therefore (since these two trajectories have the same acceleration but υ meets f tangentially at some point $\tau \leq d$) crosses f at some point before τ, contradicting the definition of γ.

Thus γ crosses δ, and therefore it intersects the shadow of f relative to g in $[c,d]$. Indeed, it must cross this shadow at two distinct points, τ_1 and τ_2, say, for it must be above the shadow where the shadow begins, and above it where γ meets g since otherwise $f^{(2)}$ and $g^{(2)}$ would intersect in $[c,d]$ ($\gamma(\tau_2)$ may be lower than $f^{(1)}(\tau_2)$ because of filling).

It is intuitively clear (and may be justified by an appeal to Lemma 7) therefore that the trail γ' of the shadow of g in $[c,d]$ relative to $f^{(1)}$ meets $f^{(1)}$ tangentially between τ_1 and τ_2. See Figure A. Suppose that σ is the point along the shadow (if it exists) where the shadow merges with a trajectory of acceleration M introduced by the 'filling' operation; then the point where γ' meets the shadow is to the left of σ, since otherwise γ' would meet f transversally at some point to the right of σ. Suppose that τ is the time, x the position, and v the velocity at this point of tangency: so it is the sole point of tangency of the trajectories $\upsilon_{(\tau,x,v)}$, which contains the shadow of g relative to $f^{(1)}$, and $\delta_{(\tau,x,v)}$, which contains the shadow of f relative to g. This implies that the trail of the shadow of f relative to g does not cross $g^{(2)}$ (transversally), and therefore does not cross $g^{(1)}$. As a consequence, $f^{(1)}$ and $f^{(2)}$ coincide; thus, the trail γ' of the shadow of g relative to $f^{(1)}$ does not cross $f^{(2)}$.

The trail γ' lies along the trajectory $\upsilon_{(\tau,x,v)}$ introduced above, which touches g in the interval and does not, of course, cross g (or $g^{(2)}$) in that interval. Since it touches but does not cross $f^{(2)}$, at a point on the shadow of f relative to g, if we can show that every point on that shadow is reachable (by an acceleration-bounded trajectory avoiding $f^{(2)}$ and $g^{(2)}$), then it will follow that every point on the shadow of g relative to $f^{(1)}$ is reachable also. The trail of the shadow of f relative to g does not cross f (or $f^{(2)}$) in $[c,d]$. Let σ' be a point where



Theorem 9, Figure A

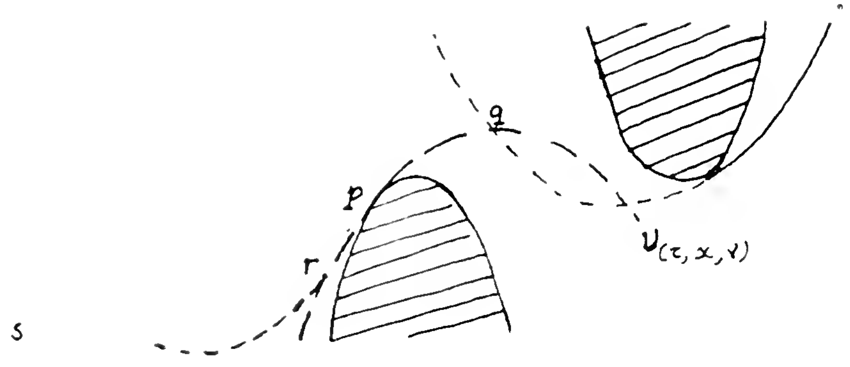
it meets g tangentially. Since the curve $\delta_{(\tau,x,v)}$ contains this trail, and touches the curve $\upsilon_{(\tau,x,v)}$ which bounds g in $[c,d]$, and the point of tangency is to the left of the point where the curve $\upsilon_{(\tau,x,v)}$ meets g in the interval, we conclude that $\sigma' < c$. By Lemma 5, the trail cannot cross $f^{(2)}$ $(= f)$ in the interval $[\sigma',c]$, and since it cannot cross $f^{(2)}$ in $[c,d]$, we conclude that it crosses neither $f^{(2)}$ nor $g^{(2)}$ in $[\sigma',d']$. By hypothesis, the point $(\sigma',g(\sigma'),\dot{g}(\sigma'))$ is reachable. This implies that all points along the shadow of f relative to $g^{(1)}$ are reachable by an acceleration-bounded trajectory which remains between $f^{(2)}$ and $g^{(2)}$, and this in turn implies that all points along the shadow of g relative to $f^{(1)}$ are reachable. Since 'filling' merely adds trajectory fragments with acceleration $\pm M$ tangent to these shadows (assuming that the original functions f and g were already filled) all points on such fragments are also reachable.

Let d' be the point where the shadow for f relative to $g^{(1)}$ begins, and let d'' be the point where the shadow for g relative to $f^{(1)}$ begins It remains to show that that part of $f^{(2)}$ with domain (c,d') and that part of $g^{(2)}$ with domain (c,d'') are reachable.

Consider a point $p = (\phi,f(\phi),\dot{f}(\phi))$, with $c<\phi<d'$. By Lemma 3 and the fact that the trajectory $\delta_{(d',f(d'),\dot{f}(d'))}$ meets the trajectory $\upsilon_{(\tau,x,v)}$ tangentially, the trajectory δ_p crosses the latter curve at some point to the right of p: let q be the earlier of the two crossing points. By a simple variant of Lemma 7, for every point r tangent to the trajectory δ_p before q, the trajectory υ_r remains underneath $\upsilon_{(\tau,x,v)}$ everywhere to the left of r. Therefore it cannot cross $g^{(2)}$ between c and r. Also, the trajectory δ_p cannot cross $g^{(2)}$ between c and q, and by definition of d' it cannot cross $g^{(2)}$ in $[a,c]$, so it cannot cross $g^{(2)}$ at any point earlier than p.

If δ_p does not cross $f^{(2)}$ at any point earlier than p then we are done. Otherwise, there is a trajectory of the form υ_r just considered which meets both it and f tangentially, at points r and s respectively, where s is before r and necessarily before c. We have seen that (if $c<r$) that part of this trajectory between c and r does not cross $g^{(2)}$; by Lemma 5, that part between s and c does not cross g $(= g^{(2)})$, so it serves to 'splice' p to a reachable point by a suitable trajectory and we conclude that p is reachable. Thus the entire segment of f between c and d' is reachable. See Figure B.

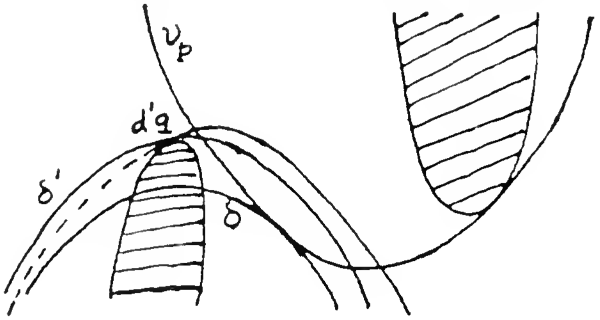
Next we consider a point p of the form $(\phi,g(\phi),\dot{g}(\phi))$ with $c<\phi<d''$. Consider the trajectory υ_p. Comparing it to $\upsilon_{(\tau,x,v)}$ and invoking Lemma 7 we see that it cannot cross $f^{(2)}$ at any point before p, and clearly it cannot cross $g^{(2)}$ before p in $[c,d]$. By Lemma 5, it cannot cross f $(= f^{(2)})$ in $[a,c]$. If it does not cross g in $[a,c]$ then we are done. Otherwise, there exists a trajectory δ which touches υ_p at r, say, and touches g at s in $[a,c]$; s is



Theorem 9, Figure B

necessarily before c. If r is before c then by Lemma 5 δ does not cross f in $[a,c]$, and we are done. So suppose that $r>c$. By Lemma 5 it does not cross f between s and c. Claim it does not cross $f^{(2)}$ between c and r. Note that if it crossed the shadow from f relative to g in $[c,d]$ then by Lemma 3 it would cross f in $[c,d]$, so it is enough to argue that it does not cross f between c and r. For suppose that this happened. Since it is above f at s and above f at r, there would be two crossing points. See Figure C. Let δ' be the (infinitely extended) trajectory obtained by translating δ upwards until it touched f at exactly one point q, say, in $[c,d]$: Lemma 3 justifies this operation, and also assures us that δ' crosses g at some point before c and crosses υ_p at some point after q. Therefore, by definition of d', q is later than d', and therefore the shadow of f against g remains above δ' after d', and therefore crosses υ_p, so p comes after d'', a contradiction. This concludes the discussion of the case where $g^{(2)} \neq g^{(1)}$.

So we only need to show the result in the case where $f^{(1)}=f^{(2)}$ and $g^{(1)}=g^{(2)}$. Suppose that the shadow is defined for g, so there is an interval $[c,d'']$ with the property that for every point $p = (\phi, g(\phi), \dot{g}(\phi))$ with $\phi \in (c,d'')$, υ_p does not cross $f^{(1)}$ to the left of ϕ. We need to construct a trajectory ending at p and not crossing the margins. Again, it may be necessary to splice υ_p to a trajectory δ with acceleration $-M$ connecting two points s and r,



Theorem 9, Figure C

one touching g at s (before c) and one touching υ_p at r. Again, we need to show that this trajectory δ does not cross $f^{(1)}$ after c, since Lemma 5 can be applied to the segment ending at c. However, if such a crossing occurred, we could again argue that δ crossed f at two points within $[c,d]$, and therefore that there was a displaced trajectory δ' touching f at some point q and crossing g in $[a,c]$. This implies that the shadow for f against g exists and begins at some point later than q. The infinite extension of δ' would necessarily cross υ_p, so therefore so would the shadow of f against g, and therefore the trail of the shadow for g against f would cross $f^{(1)}$ transversally, so $g^{(1)} \neq g^{(2)}$, a contradiction.

Thus every point in (c,d'') along the curve g is reachable. Consider the trail γ of the shadow for g against f: it touches g at d'', and f at some earlier point. If it touches f in $[a,c]$ then we know by previous arguments that it is reachable. Otherwise, it touches f at some point e in (c,d''). If the shadow is also defined for f, then we know that the point d' where it meets f is no earlier than e. If $e<d'$, then we already know that f is reachable at e, so the shadow to g is reachable; otherwise, $e=d'$, but in this case, the trail of the shadow for f against g must meet g at some point earlier than d'', i.e., at a reachable point, so the shadow for f against g is reachable at d', and consequently the shadow for g against f is reachable in this case as well. This concludes the proof. **Q.E.D.**

6. Algorithmic consequences.

The results of the last two sections lead us to an algorithmic solution of the original problem, for the following reasons. We assume a computational model in which square-root extraction takes $O(1)$ operations and is exact. Let n be the total number of times that $\ddot{f}$ and $\ddot{g}$ change.

In order to automatically satisfy the initial conditions specified by (2.0), we choose the trajectories δ and υ with accelerations $-M$ and M respectively beginning at (a,x_a,v_a), splice υ to g (by 'filling') and δ to f. This adds at most four new segments in which $\ddot{f}$ or $\ddot{g}$ changes.

We compute the reachable margins $(\hat{f},\hat{g})$ of f and g by iterating the construction described in the previous section. First observe that the process of 'filling' f and g introduces at most n new intervals where $\ddot{f}$ or $\ddot{g}$ changes, since each interval in which f or g is simple is incident to at most two segments introduced by filling. Also, the process of 'shading' introduces at most n new intervals, since each 'shadow' of f (respectively, g) meets f (g) at a distinct interval in which $\ddot{f}<-M$ (respectively, $\ddot{g}>M$). Therefore the complexity of $\hat{f}$ and $\hat{g}$ is linear in n.

Given g, it is possible to construct the filled version $\bar{g}$ in time $O(n^2)$. To do this one processes g from left to right. Suppose that $[c,d]$ is an interval in which g is simple. Suppose, also, that $\ddot{g} \geq -M$ on $[a,c]$. If $\ddot{g} < -M$ on $[c,d]$, locate in time $O(n)$ (by naive methods), the latest point e in $[c,d]$ such that there exists a trajectory δ with acceleration $-M$ touching g at e and also at some earlier point σ: such a trajectory must exist, perhaps touching G only at the 'corner' at a; then replace g by δ in the interval $[\sigma,e]$. If $\ddot{g} < -M$ it is unnecessary to do anything unless $d = b$ in which case one finds the appropriate trajectory δ passing through $g(b)$ at b. (This is not optimal: with some care one can construct $\bar{g}$ in time $O(n)$.)

Computing the shadows of f and g can be done similarly in time $O(n)$ by naive methods, and having done so, one can splice them to the appropriate 'filling' trajectories by naive methods in time $O(n)$.

It is straightforward to verify in time $O(n)$ whether f and g cross by a left-to-right scan. Finally, observe that while iterating the construction may introduce new points where $\ddot{f}$ or $\ddot{g}$ changes, these points are always later in time so the processing can continue from left to right.

Thus supposing that $\hat{f}$ and $\hat{g}$ have been constructed, given the final position and velocity (x_b, v_b), one can determine whether it is reachable by computing $v_\ell(x_b)$ and $v_u(x_b)$ and determining whether v_b lies between them. (It is necessary, of course, that $\hat{f}(b) \leq x_b \leq \hat{g}(b)$.) This can be done naively in time $O(n)$.

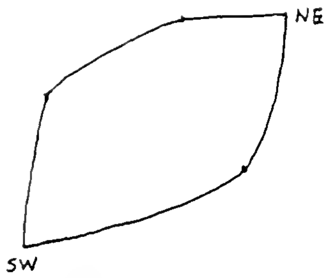
Although this solves only the existence problem, assuming that the point is reachable, an admissible trajectory reaching it can be constructed in $O(n^2)$ time by iterative application of the method of Theorem 6.

7. Standard phase-regions.

A *standard phase region* I is a bounded, convex region in phase space, expressible in the form

$$\{(x,v):\ \phi_1(v) \leq x \leq \phi_2(v)\}$$

where ϕ_i are continuous, piecewise linear/quadratic, nondecreasing functions of v defined on some common domain $[v_{min}, v_{max}]$. Furthermore, the function ϕ_1 is strictly increasing except that it may be constant on some terminal segment $[w, v_{max}]$, and ϕ_2 is strictly increasing except that it may be constant on some initial segment.



Illustrating a standard phase region.

Lemma 10. (i) If I is a standard phase region at time a, then $FR_{a,b}(I)$ is a standard phase region at time b. (ii) If f and g form a reachable pair of pursuit functions over an interval $[a,b]$, then the phase region they form at time b is a standard phase region.

Proof. (i) We know that $FR_{a,b}(I)$ can be expressed as a Minkowski sum

$$T_{a,b}(I) + FR_{a,b}(0,0)$$

where $T_{a,b}$ is the linear transformation taking (x,v) to $(x+(b-a)v,v)$ and $FR_{a,b}(0,0)$ is a region bounded by two segments definable by quadratic functions of v, each being monotonic over their common domain of definition. Clearly $T_{a,b}(I)$ is a standard region. It is a straightforward exercise to verify that the Minkowski sum of these two regions is again a standard region.

(ii) We know that the region they form at b can be defined as

$$\{(x,v) \colon x \in [f(b),g(b)] \text{ and } v \in [v_\ell(x),v_u(x)]\}$$

where v_ℓ and v_u are defined as in Section 4. Let us concentrate on $v_\ell(x)$. By Lemma 3, we know that if $x<x'<g(b)$ and $v = v_\ell(x)$, then the trajectory $\delta_{(b,x',v)}$ meets G transversally, so $v < v_\ell(x')$. This proves that v_ℓ is strictly increasing on $[f(b),g(b))$, so it has a well-defined inverse which we shall call ϕ_1 in this interval. Suppose that ϕ_1 is defined over the interval $[v_{min},w)$. By the analysis in Appendix B, it is piecewise linear/quadratic and strictly increasing but with nonpositive second derivative in that interval, and $u = v_\ell(g(b))$. The function ϕ_1 can be extended to be constant in $[w,v_u(g(b))]$. This shows that ϕ_1 has the required format. The rest of the analysis follows by considerations of symmetry. Q.E.D.

In fact, if I is a standard phase region at time a then it is relatively trivial to compute $FR_{a,b}(I)$, because of the following observation.

Lemma 11. Let U be the upper boundary of a standard phase-region at time a, and let U' be the curve

$$T_{a,b}(U) + (-\tfrac{1}{2}M(b-a)^2, -M(b-a))$$

(using the notation of Minkowski sum). Then the upper boundary of $FR_{a,b}(I)$ is obtained by splicing U' to a copy of the upper boundary of $FR_{a,b}(0,0)$.

Proof. Let X represent the upper boundary of $FR_{a,b}(0,0)$. The significant point is that the upper boundary of $FR_{a,b}(I)$ is the upper boundary of the Minkowski sum $U' + X$. The slope of X is maximal at its base $(-\tfrac{1}{2}M(b-a)^2, -M(b-a))$: a brief calculation shows that its slope at that point is $b-a$. The slope of U' is minimal at its summit, at which its value is

$$\frac{d}{dv}(\phi_1(v) + (b-a)v) \text{ at } v = v_{max},$$

i.e., at least $b-a$. This implies that the upper boundary, which is the envelope of curves of the form

$$\{(x,v) + X : (x,v) \in U'\}$$

is swept out by the base of X, and the result follows. •

For computational purposes, the upper and lower boundaries of I can each be represented as a sorted sequence of parabolic and linear segments. Notice that the linear transformation $T_{a,b}$ takes a parabolic segment to another with the same second derivative but displaced apex, and, since $T_{a,b}$ is invertible, the number of segments is unchanged, and, since the upper and lower boundaries are monotonic, $T_{a,b}$ preserves the order of these segments. In other words, a careful representation will enable $FR_{a,b}(I)$ to be maintained 'generic' for all $b > a$.

For any point $p = (\tau, x, v)$ in phase space at time τ, let

$$FR_{a,\tau}^{-1}(p) = \{(a, x_a, v_a) : p \in FR_{a,\tau}(a, x_a, v_a)\}. \quad (7.1)$$

so it is the set of points q in phase space at time a from which p can be reached by an

U and U'

acceleration-bounded trajectory.

Again by abuse of notation, we shall often ignore the first coordinate in this set of triples, and simply regard it as a subset of phase space (for example, in the following identity). One can show by the methods of Section 3 or Section 4 that

$$FR_{a,\tau}^{-1}(\tau,x,v) = (x-(\tau-a)v,v)+FR_{0,\tau-a}^{-1}(0,0),$$

and $FR_{0,t}^{-1}(0,0)$ is a convex region in phase space bounded by two parabolas with apexes $A = (\frac{1}{2}Mt^2,-Mt)$ and $B = (-\frac{1}{2}Mt^2,Mt)$ and second derivatives $\frac{-1}{2M}$ and $\frac{1}{2M}$ respectively; the points A and B are the points of intersection of the two parabolas.

Lemma 12 below provides some technical details useful for this section. Its proof is deferred to Appendix C. The following notation is used. The upper pursuit function g is supposed simple over an interval $[c,d]$, and its acceleration N in that interval is assumed to be greater than $-M$. We write $X(\tau)$ to abbreviate $FR_{c,\tau}^{-1}(\tau,g(\tau),\dot{g}(\tau))$. Let $Y_1(\tau)$ represent the set of all points in phase-space lying in or above $X(\tau)$:

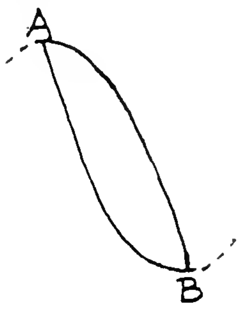
$$Y_1(\tau) = \{(x,v)\colon \exists x'\leq x((x',v)\in X(\tau))\}, \tag{7.2}$$

and $Y_2(\tau)$ be the set of points lying in $Y_1(\tau)$ or northwest of the northwest corner A of $X(\tau)$:

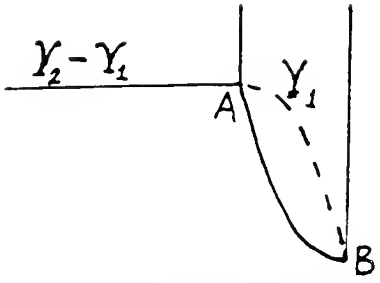
$$Y_2(\tau) = Y_1(\tau)\cup\{(x,v)\colon r\geq x \text{ and } s\leq v\}, \tag{7.3}$$

where $(r,s) = A$ is the northwest corner of $X(\tau)$. In the statement of the following Lemma, '$Y(\tau)$ increases strictly with τ' means that given $\tau_1<\tau_2$ in the domain of definition, $Y(\tau_1)\subseteq\text{Interior}(Y(\tau_2))$.

Lemma 12. With the notation introduced above, (i) if $-M<N<M$, then $Y_1(\tau)$ increases strictly with τ; (ii) if $N=M$ then $X(\tau)$ increases strictly with τ along the lower



Illustrating $FR_{0,t}^{-1}(0,0)$.

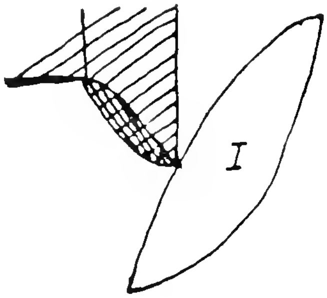


Illustrating Y_1 and Y_2.

boundary whereas its upper boundary remains within the same parabolic curve; (iii) if $N>M$, then $Y_2(\tau)$ increases strictly with τ.

Lemma 13. Suppose that I is a standard phase-region at time c, and g is an upper pursuit function such that g is simple and $\ddot{g}>-M$ on $[c,d]$. Then the set of points along g reachable from I, i.e., the set of points $p = (\tau, g(\tau), \dot{g}(\tau))$ for which there exists an acceleration-bounded trajectory beginning at I and ending at p and not crossing g, is either empty or a connected subinterval $[c', d']$ of $[c, d]$.

Proof. Clearly, a point $(\tau, g(\tau), \dot{g}(\tau))$ is reachable if and only if $X(\tau)$ intersects I. It is not difficult to show by continuity arguments that the set of such points τ is closed. Let K be a connected component of this set, assuming it to be nonempty: write $K = [c', d']$. Clearly, if $c' \neq c$ then $X(c')$ meets I tangentially. First observe that the point of tangency cannot be outside the lower boundary of $X(\tau)$ and the upper boundary of I: for although it is possible (if $N<M$) that the upper boundary of $X(\tau)$ could begin to penetrate I first, in this case, at the first point where the sets met tangentially, there would be a unique trajectory from I to $(\tau, g(\tau), \dot{g}(\tau))$, and it would approach $g(\tau)$ from above and therefore not be admissible. Thus we can assume that for $\tau = c'$ the point of tangency is along the lower boundary of $X(c')$ and the upper boundary of I (possibly at its northeast corner). Clearly in this situation



Illustrating the argument in Lemma 13.

both $Y_1(c')$ and $Y_2(c')$ meet I tangentially. Since one of these families of sets (depending on whether $N>M$ or $N\leq M$) grows strictly with increasing τ, we conclude that c' is unique; hence K is unique and defines the entire set of points reachable along g.

Note that if $d'<d$ then $N>M$ and the upper boundary of $X(d')$ meets the lower boundary of I tangentially. •

In the above lemma it is possible that the set $X(d)$ could not meet I at all but $g(d)$ still be reachable by some trajectory whose velocity was strictly greater than $\dot{g}$ at d. For consider the family of sets

$$Z(v) = FR_{\dot{c},d}^{-1}(d,g(d),v) \quad \dot{g}(d)\leq v<\infty .$$

It is easy to see that $Z(\dot{g}(d)) = X(d)$, and in general as v increases the sets $Z(v)$ sweep downwards with their centers along the line $(g(d)-(d-c)v,v)$. If we define $Y_2(v)$ analogous to the definition given for $Y_2(\tau)$, in the obvious way, it is easy to show that these sets increase strictly with v, so there may be a minimal $v>\dot{g}(d)$ such that $Z(v)$ is tangent to I: if so, this v is unique, and there is a unique trajectory connecting I to $(d,g(d),v)$. In this case, we shall regard the set K of Lemma 13 as defining a degenerate interval of reachable points with $c' = d$.

Continuing this notation, if $c'=c$ then g is reachable from I at c; otherwise (assuming that g is reachable somewhere in $[c,d]$), there is a unique trajectory beginning in I and touching g at c'. Let us call this trajectory the *connection* between I and g in $[c,d]$. Again, if $d'<d$ there is a unique trajectory from I touching g at d'. In this case, $M<N$: we call the trajectory $\upsilon_{(d',g(d'),\dot{g}(d'))}$ restricted to $[d',d]$ the *shadow* from g against I in $[c,d]$, and the unique trajectory from I ending at d' the *trail* of the shadow.

Let us consider briefly the possibility that the set of points discussed in Lemma 13 be empty. In particular, there is no acceleration-bounded trajectory beginning in I and ending at $g(d)$ but not crossing g. Note that the existence of at least *one* trajectory beginning at I and not crossing g is easily tested, for it exists if and only if the lowest possible trajectory, i.e., that beginning at the southwest corner of I and maintaining acceleration $-M$, avoids crossing g.

Suppose that the lowest trajectory has this property, so there is a well-defined point (d,x_h) reachable by some (accleration-bounded) trajectory from I which does not cross g, such that x_h is maximal. Let γ be some such trajectory from I which reaches this point. Suppose that $x_h<g(d)$. Clearly, if γ touches g at some point σ but not later, where $\sigma<d$, then $\ddot{\gamma}=M$ in $[\sigma,d]$. In particular, if $\sigma>c$ then $\ddot{g}>M$ and γ defines a shadow from g

against I, and the interval $[c',d']$ of Lemma 13 is not empty.

If σ is defined and equal to c then clearly γ is the trajectory beginning at $(c, g(c), w)$ in I where w is maximal; again, we regard the trajectory γ as the shadow from I against g in $[c,d]$.

Otherwise γ meets g nowhere in $[c,d]$; suppose that it originates from some point (c, x, v) in I. If this point is not the northeast corner of I then by perturbation arguments $\gamma(d) = x_h$ is not maximal, a contradiction: therefore, γ is the trajectory leaving this corner at acceleration M, defines the upper boundary of all points reachable from I, remains below g in $(c,d]$, and is unique: again, we regard this as defining the 'shadow' of g against I in $[c,d]$.

Let us extend this analysis to cover the case where $\ddot{g} < -M$ in $[c,d]$. If $g(d)$ is not reachable, the previous analysis applies and we call the highest trajectory from I which does not cross g the shadow; if $g(d)$ is reachable, then there are two possibilities: (i) the trajectory of acceleration $-M$ meeting g at c and d (there is exactly one such trajectory; call it δ) originates in I; in this case we call this trajectory the connection to I; (ii) the trajectory δ originates at some point outside I; then the arguments about the sets $Z(v)$ apply, and there is again a unique trajectory originating in the boundary of I and meeting g with minimal velocity at c; this again we call the connection between I and g.

Notice that with these definitions the connection and shadow of g, where defined, are reachable throughout by trajectories which do not cross g and originate in I.

Lemma 14. If $\bar{g}$ is obtained by splicing g to its connection and shadow (where defined), then no point above $\bar{g}$ lies in any acceleration-bounded trajectory beginning in I and remaining on or below g throughout $[c,d]$.

Proof. First suppose that the set $[c',d']$ defined in Lemma 13 (and the remarks following it) is nonempty. It is easy to see that a connection to g begins with a (possibly empty) curve υ of acceleration M followed by a curve δ of acceleration $-M$. Consider any acceleration-bounded trajectory passing above δ at some point (τ, x, v) say. Let $v' = \dot{\delta}(\tau)$. By Lemma 3 $v < v'$ since otherwise the given trajectory would cross g in $[c,d]$. Let B be the southeast corner of $FR_{c,\tau}^{-1}(\tau, x, v)$. Then this point is northwest of the point where the connection begins, and, by the arguments in Lemma 13, cannot meet I.

If (τ, x, v) is a point above the shadow of g, and we have an acceleration-bounded trajectory reaching it, that trajectory must either cross g or cross the shadow, and in the latter case by similar arguments it cannot originate in I.

In the other cases discussed above, where no part of g is reachable from I, the analysis is obtained by direct arguments. ●

The following construction produces what we shall call the 'margins' of f and g (relative to I).

Compute the connections and shadows for g and f; Let $f^{(1)}$ and $g^{(1)}$ be obtained by splicing f and g to these extra trajectories where applicable.

Compute the shadow from g against $f^{(1)}$ and the shadow from f against $g^{(1)}$ in the interval $[c,d]$ (in the sense of Section 5). If the shadow of g against $f^{(1)}$ is lower than $g^{(1)}$ on its domain of definition, splice it to g; similarly for f. The resulting functions $f^{(2)}$ and $g^{(2)}$ are called the ***reachable margins*** of f and g from I in $[c,d]$.

By Lemma 14 (and Lemma 8), if these margins cross, then there exists no acceleration-bounded trajectory beginning at I and not crossing f or g in the interval. On the other hand, we can now show that if the margins do not cross, then all points along these margins are reachable in the sense that there always exists a trajectory from I ending tangential to the margin at any point and not crossing the margins before that point.

Lemma 15. If these margins do not cross, then all points along them are reachable.

Proof. Note that in constructing the margins it was ensured that $(c, g^{(2)}(c), \dot{g}^{(2)}(c))$ belonged to I. Therefore, if $\ddot{g}^{(2)} \leq M$ in the interval $[c,d]$, then all points along it are reachable. Suppose that $\ddot{g} > M$, and we know already that over some subinterval $[c,e]$ where $c \leq e \leq d$ all points along both $f^{(2)}$ and $g^{(2)}$ are reachable. Certainly we can assume that e is not strictly within the domain of the connection to I, since that curve is automatically reachable and if $\ddot{f} < -M$ and the connection for f is earlier then we can treat f in place of g. If $e=d$ or e is the point where a shadow meets g, then we know that all later points on the upper margin are reachable. So suppose that e is before a shadow to g (so implicitly $\ddot{g} > M$). Let $e' > e$ be chosen so small that in the interval (e,e'), the domain of the shadow (if defined) does not include e' and the trajectories $\delta_{(\tau, g(\tau), \dot{g}(\tau))}$ and $\upsilon_{(\tau, g(\tau), \dot{g}(\tau))}$ do not cross $f^{(2)}$ or $g^{(2)}$ in the interval (e,e') for any τ in (e,e').

We may restrict our attention to the upper margin. Let τ be any point in (e,e'): we want to show that the point q is reachable where $q = (\tau, g(\tau), \dot{g}(\tau))$. Consider the trajectory υ_q: by hypothesis it does not cross $f^{(1)}$ in $[c,\tau)$. Claim that it cannot cross $f^{(2)}$ in that interval. For otherwise, $f^{(2)} \neq f^{(1)}$, and the shadow for f against $g^{(1)}$ is defined and begins at some point d', say. By the usual arguments (see Theorem 9), υ_q crosses this shadow twice in $[d',\tau)$. Thus the trail of this shadow must meet $g^{(1)}$ at some point where $\dot{g}^{(1)} = +M$ (it

must meet $g^{(1)}$ left of q). In particular it is on the connection to I which begins at the northeast corner of I. See figure A. This implies that the shadow for f against $g^{(1)}$ coincides with the shadow for f against I and $f^{(1)}=f^{(2)}$, a contradiction. Thus υ_q does not cross $f^{(2)}$.

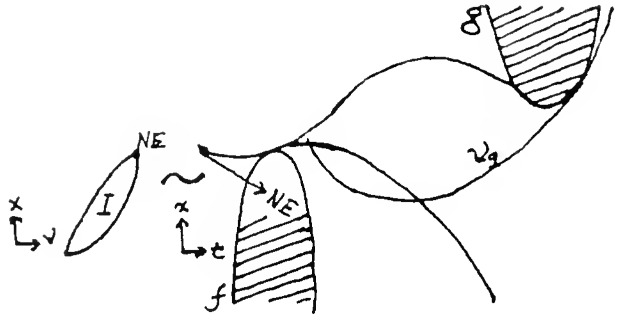
Let $A = (\upsilon_q(c), \dot{\upsilon}_q(c))$. If (i) $A \in I$ and (ii) υ_q does not cross $g^{(2)}$ in $[c,\tau)$ then we are done. Otherwise suppose first that υ_q does not cross $g^{(2)}$ but A is not in I. Let

$$X'(\sigma) = FR_{c,\sigma}^{-1}(\sigma, \upsilon_q(\sigma), \dot{\upsilon}_q(\sigma)).$$

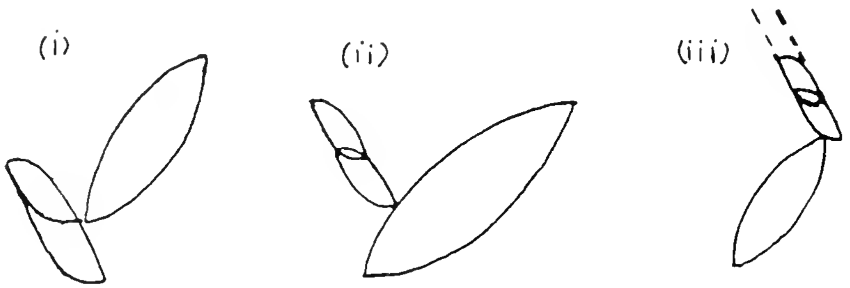
for $\sigma \in [c,\tau]$. Thus $X'(c) = \{A\}$ and $X'(\tau) = X(\tau)$ which by hypothesis meets I. Let σ be the (unique) time such that $X'(\sigma)$ meets I tangentially. There are several possibilities for this tangency. See Figure B.

(i) The upper boundary of $X'(\sigma)$ meets the southwest corner of I. In this case, since the upper boundary of $X'(\sigma)$ is always part of the upper boundary of $X(\tau)$ (Lemma 12), we conclude that the same holds for $X(\tau)$. Therefore it is the southeast corner of $X'(\sigma)$ which meets the southwest corner of I, and case (ii) applies.

(ii) The southeast corner B of $X'(\sigma)$ meets the upper boundary of I. In this situation, there is a unique trajectory δ from B meeting υ_q at σ, of acceleration $-M$. We need to argue that δ does not cross $f^{(2)}$. Otherwise, since $\delta(\sigma) = \upsilon_q(\sigma) > f^{(2)}(\sigma)$, δ crosses $f^{(2)}$ at



Lemma 15, figure A.



Lemma 15, figure B.

one or two points in $[c,d]$. Again we note that δ crosses f at some point q' with velocity less than $\dot{f}$ at that point, and we deduce by the previous arguments that f is not reachable from I at that point, since the set of points (c,x,v) from which f is reachable at that point would lie northwest of I, a contradiction: therefore δ does not cross $f^{(2)}$ and it can be used to complete a trajectory reaching g at q.

(iii) The lower boundary B of $X'(\sigma)$ meets the northeast corner of I. Again, there is a unique (acceleration-bounded) trajectory connecting B to υ_q at σ, and again we can argue that if it crosses $f^{(2)}$ then there is some earlier point along f which is not reachable from I. For suppose that it crosses $f^{(2)}$, and therefore f, at some point σ': if there are two such crossing times choose the later one; suppose that x is the position and v the velocity at this crossing, so $x = f(\sigma')$ and $v > v' = \dot{f}(\sigma')$. Consider the family of sets

$$Z(w) = FR_{c,\sigma'}{}^{-1}(\sigma',x,w), \quad -\infty < w \leq v.$$

It is a straightforward exercise (following the methods of Lemma 11 and the observation that the sets $Z(w)$ are being swept along a line with slope $c-\sigma'$) to show that the region swept out by these sets $Z(w)$ is bounded by the lower boundary of $Z(v)$ and two straight lines at slope $c-\sigma'$ through the two corners of $Z(v)$. Therefore $Z(v')$ does not meet I, and f is not reachable at σ', a contradiction.

This covers the case where υ_q does not cross $g^{(2)}$ in $[c,\tau]$. Suppose, then, that it does; let δ be a 'splicing' trajectory (of acceleration $-M$) connecting υ_q to $g^{(2)}$. By the previous arguments, δ cannot cross $f^{(2)}$. If it meets $g^{(2)}$ (tangentially) at some point σ later than c, then that is earlier than e and again by hypothesis that point is reachable and hence so is g at q.

Suppose that δ meets $g^{(2)}$ at c with a velocity $\dot{\delta} < \dot{g}$; if B represents its position and velocity at c then B, being west of $(g(c),\dot{g}(c))$, is west of I or in I, and the above arguments may again be applied to find a trajectory from I meeting υ_q tangentially and not crossing $f^{(2)}$ nor $g^{(2)}$. **Q.E.D.**

Next we assume that the above procedure has been applied so f and g are pursuit functions of bounded complexity (at most four changes of acceleration) in $[c,d]$, and every point along f and g is reachable by a trajectory originating in I and not crossing f or g at any earlier point. In this case we say that f and g form a *reachable pair relative to I*

Theorem 16. Suppose that I, f, and g are given, I being a phase region at time c, and f and g forming a reachable pair relative to I, being defined on the interval $[c,d]$. Let S denote the phase region at time d formed from f and g. Then

$$R_{c,d}(I) = S \cap FR_{c,d}(I).$$

Proof. For any x in $[f(d),g(d)]$, let $V_\ell(x)$ be the minimum velocity such that the trajectory δ of acceleration $-M$ ending at (d,x,v) meets but does not cross g. By Lemma 4 we know that for any $v<V_\ell(x)$ there is no admissible trajectory ending at (d,x,v). If δ meets g tangentially at some point then by the reachability hypothesis all points along δ are reachable (note that δ cannot cross f by the usual arguments). In this case, $(d,x,V_\ell(x))$ is in the boundary of S. Otherwise it meets g at c. Let $B = (\delta(c),\dot\delta(c))$ be the point in phase space where δ crosses c. Then since $(g(c),\dot g(c))$ is in I, B is either in I or west of I. If $B \in I$ then again $(d,x,V_\ell(x))$ is in $FR_{c,d}(I)$ and on the boundary of S. Otherwise B is west of the upper boundary of I. Consider the family of sets

$$Z(v) = FR_{c,d}{}^{-1}(d,x,v), \quad v \geq V_\ell(x).$$

Claim that there exists another $v>V_\ell(x)$ such that $Z(v)$ meets I tangentially. To show this, we define $V_u(x)$ and a trajectory υ meeting f in the obvious analogous fashion. By parallel reasoning, either υ meets f tangentially, so $Z(V_u(x))$ intersects I (by the reachability hypothesis) or the northwest corner of this set is either in or east of I. In any event, the set swept out by $Z(v)$ as v varies over $[V_\ell(x),V_u(x)]$ must intersect I and there exists an earliest velocity $v_\ell(x)$ such that $Z(v_\ell(x))$ meets I tangentially.

Following the arguments in Lemma 15, we know that the region swept out by the sets $Z(v)$ is bounded above by a straight line through the B at (negative) slope $c-d$, and therefore they are bounded away from the highest part of I. Therefore the southeast corner of of $Z(v_\ell(x))$ meets I tangentially, and there is a unique trajectory, call it δ', originating at this point and passing through $(d,x,v_\ell(x))$, with acceleration $-M$. By the usual reasoning this trajectory cannot cross f since otherwise not all of f would be reachable, and it cannot cross g since $v_\ell(x)>V_\ell(x)$. Note that $v_\ell(x)\leq V_u(x)$ and $(d,x,v_\ell(x))$ is on the boundary of $FR_{c,d}(I)$.

Thus in any case we have established a lower bound $v_\ell(x)$ on the range of velocities v such that (d,x,v) is reachable by an admissible trajectory from I, and that $(d,x,v_\ell(x))$ is indeed reachable, and it belongs to the boundary of $S \cap FR_{c,d}(I)$. Similarly one can construct an upper bound $v_u(x)$.

Now choose any v in the range $(v_\ell(x),v_u(x))$; we want to argue that (d,x,v) is reachable by an admissible trajectory originating in I. Let υ be the trajectory with acceleration M ending at (d,x,v). By definition of $v_u(x)$, υ does not cross f. If υ does not cross g and

originates in I then it is the kind of trajectory we want and there is nothing more to show. Suppose, then, that υ does not cross g but originates from outside I. Note that the point $A = (\upsilon(c), \dot{\upsilon}(c))$ where υ crosses c is the northwest corner of the set $Z(v)$ defined above, and by hypothesis this set intersects I.

Therefore this point A is west of I. For $\sigma \in [c,d]$, define

$$X'(\sigma) = FR_{c,\sigma}^{-1}(\sigma, \upsilon(\sigma), \dot{\upsilon}(\sigma)).$$

Thus $X'(c) = \{A\}$ and $X'(d) = Z(v)$. Therefore there exists an intermediate point σ such that $X'(\sigma)$ meets I tangentially; and by repeating the previous arguments we know that the tangency is between the southeast corner B of $X'(\sigma)$ and the (western) boundary of I. This implies that there is a unique trajectory δ which begins at B and meets υ tangentially at σ. Since δ is below υ, it does not cross g; since it begins and ends above f, by the reachability hypothesis it cannot cross f; this shows that (d,x,v) is reachable by an admissible trajectory.

Therefore the one case left to cover is where υ crosses g. Let δ' be a trajectory meeting but not crossing g and meeting υ tangentially, at some point σ'. By construction δ' does not cross g and by the reachability hypothesis it does not cross f. If $\sigma' > c$ or $\sigma' = c$ but δ' crosses c at some point in I then previous arguments apply. Otherwise the point B at which δ' crosses c, the southeast corner of $X'(\sigma')$, must be west of I, and we can continue the same arguments as before to find a later point σ such that $X'(\sigma)$ meets I tangentially, and continue as before.

We have shown that for any x in the range $[f(d), g(d)]$, the set K of velocities v such that (d,x,v) is reachable by an admissible trajectory originating in I is a connected interval whose endpoints are on the boundary of S or of $FR_{c,d}(I)$. Clearly, $K \subseteq S \cap FR_{c,d}(I)$. It follows easily that K coincides with the cross-section of $S \cap FR_{c,d}(I)$ at x. Thus $R_{c,d}(I) = S \cap FR_{c,d}(I)$. **Q.E.D.**

Continuing the notation of Theorem 16, consider the upper boundary of $S \cap FR_{c,d}(I)$. Clearly it intersects the upper boundary of S (assuming f and g form a reachable pair relative to I). It may or may not intersect the boundary of $FR_{c,d}(I)$ (implicitly in the proof of Theorem 16, at a point V in the upper boundary). Let $W = (\delta(d), \dot{\delta}(d))$ be the point where the trajectory δ originating at $(g(c), \dot{g}(c))$ with acceleration $-M$ crosses d. If $\delta(d) < f(d)$ then no such point V exists. Otherwise, we know (see appendix B) that the upper boundary of S has a linear segment beginning at W with slope $d-c$, so V is where this line-segment first intersects the upper boundary of $FR_{c,d}(I)$, if it does. Since W is in $FR_{c,d}(I)$, V, if it exists, is unique, and the upper boundary of $R_{c,d}(I)$ is obtained by splicing the upper

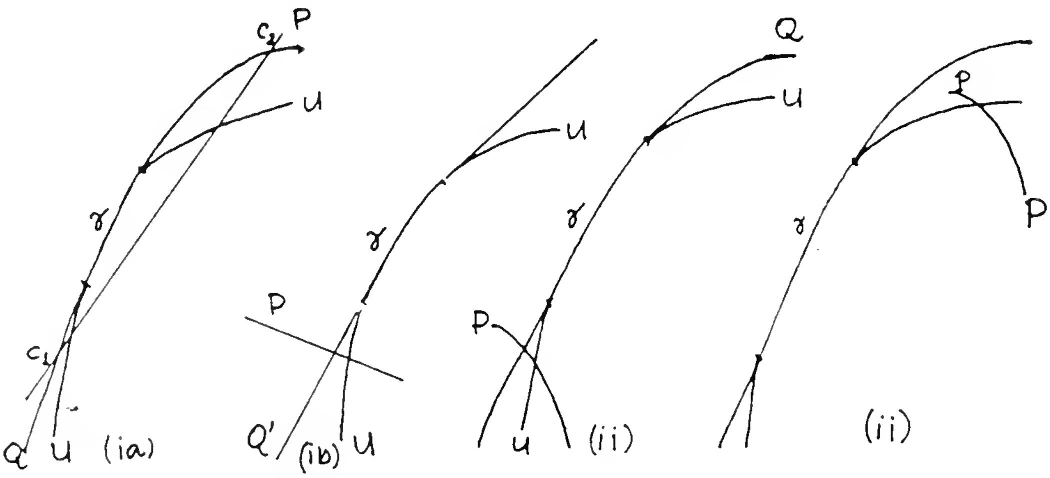
boundary of $FR_{c,d}(I)$ to the upper boundary of S at this point, and truncating it by the line $x=f(d)$ if necessary.

The preceding analyses provide a basis for a more efficient $O(n\log(n))$ solution to the reachability problem. We separate $[a,b]$ into intervals with respect to which both pursuit functions are simple. Beginning with $I = \{(a,x_a,v_a)\}$ we compute iteratively the sets reachable at the intermediate points. They are all standard regions and their boundaries can be represented as two chains of linear/quadratic functions, meeting at intermediate nodes. They can be maintained in sorted order with suitable usage of balanced trees.

The critical operations involved are (i) to compute the margins as described in Lemma 15 and (ii) to compute the boundary of $S \cap FR_{c,d}(I)$ (in the notation of Theorem 16).

Lemma 17. Suppose that U represents the upper boundary of a standard phase-region I (or of $FR_{c,d}(I)$), and P is either (i) an infinite straight line, (ii) part of a quadratic curve with negative second derivative, bounded on the left by a point at or to the right of its apex, or (iii) part of a quadratic curve with positive second derivative, bounded on the right by its apex. Then assuming U to be represented by a sorted list of its segments and endpoints of size $O(n)$, it is possible to determine in $\log(n)$ time whether U intersects P, and, if so, where.

Proof. For convenience we assume that U is extended infinitely in both directions by adding the tangents to it at its lower and upper ends: let θ_1 and θ_2 be the slopes of these lower and upper infinite segments respectively (oriented from their points of origin so θ_1 is negative and θ_2 positive), let γ be the median segment of U, and Q the infinite linear or semi-infinite quadratic (bounded on right by its apex) extension of γ.



Illustrating various cases in Lemma 17

Case (i). If (a) θ_1 and θ_2 are both on the same side of P then there can be two points of intersection, else (b) there must be exactly one.

Case (ia). If P does not meet Q then it does not meet U, and if it touches Q then it touches U if and only if the point of tangency is in γ; otherwise, let c_1 and c_2 be the two points where the curves intersect (if there is only one replace c_2 by the apex of Q). Without significant loss of generality we are searching for the lower of the two possible intersections of U with P. This lower point is in or before γ (along U) if and only if c_1 is in or before γ (along Q), reducing the search space by half; thus $O(\log(n))$ probes suffice.

Case (ib). For this case, it is convenient to consider not Q but (when γ is nonlinear) the curve Q' defined by extending γ by the infinite tangents to it at its endpoints. Then there is a unique point c where P meets Q', and again the point where P meets U is before, in, or after γ along U if and only if c is respectively before, in, or after γ along Q'. Thus again, $O(\log(n))$ probes suffice.

Case (ii). In this case, P can meet Q at most once. If P meets Q at a point c, then P meets U and it meets U before, in, or after γ if and only if it meets Q before, in, or after γ. If P does not meet Q, consider where its apex p is placed relative to γ. If p is higher (in the x-direction) than the upper endpoint of γ then P can meet U only after γ; if it is lower than the lower endpoint of γ then P can meet U only before γ; otherwise P cannot meet U. Thus $O(\log(n))$ probes suffice.

Case (iii). In this case it is more convenient to consider the curve Q' of case (ib). If P does not meet Q' then it does not meet U; otherwise, it can meet U before/in/after γ if and only if it meets Q' before/in/after γ. Thus $O(\log(n))$ probes suffice. •

Suppose that a standard region I at time c is given, and f and g are simple pursuit functions over the interval. The observations in Lemma 17, taken with the construction in Lemma 13 and the discussion of exceptional cases following Lemma 13, allow us to compute in $O(\log(n))$ time the connection and shadow to g, where defined. We assume that the upper boundary of I is maintained in a suitable balanced-tree representation, which is generic so that (assuming that I is a standard phase-region at time c), queries of the above form can be answered in logarithmic time for a set $FR_{c,d}(I)$, for any $d \geq c$ (this is possible from the discussion immediately following Lemma 11). Similarly another balanced-tree generic representation is assumed for the lower boundary of I. Typical steps in the computation are as follows.

(A) Is $(g(c), \dot{g}(c))$ in I? Use the methods in Lemma 17 to compute where the horizontal line through this point intersects I, if at all, and then ascertain if the given point is

contained in this intersection.

(B) If $\ddot{g}>-M$, to compute the 'connection,' if any. Query (A) tells us whether the connection is nontrivial. If it is (so $X(c)$ does not meet I in the notation of Lemma 13), one can ascertain in constant time whether there exists a τ such that $X(\tau)$ meets the northeast corner of I. If not, it remains to ascertain whether there exists a τ such that the southeast corner $B(\tau)$ of $X(\tau)$ meets the upper boundary of I. But (as in Appendix C) we know that this corner traces a parabolic curve P as in case (ii) of Lemma 17, and hence the query can be answered in $O(\log(n))$ time.

(C) For fixed τ, to ascertain whether $X(\tau)$ meets I. In constant time, one can ascertain whether the line-segment L connecting the extreme corners of I intersects $X(\tau)$. If not, then $X(\tau)$ intersects I if and only if its boundary (a) intersects the boundary of I or (b) is contained entirely within I. Both these cases can be handled by considering the two semi-infinite parabolas containing the boundary of $X(\tau)$ and finding where they intersect the boundary of I.

(D) If $\ddot{g}>M$ and $X(d)$ does not intersect I, one can ascertain whether there exists a d' before d such that $X(d')$ is tangent to I by methods similar to those in (B).

(E) Given the hypotheses of Theorem 16, and following the notation of that theorem, we want to ascertain where the upper boundary of S meets the upper boundary U of $FR_{c,d}(I)$. This involves locating where, if ever, a given line of slope $d-c$ intersects U, a problem solvable by the methods of Lemma 17.

8. Concluding remarks.

This paper has given polynomial-time algorithms for the admittedly narrow problem set out in the Introduction. There are two obvious limitations: (a) the problem is strictly one-dimensional and (b) the pursuit functions are restricted to be quadratic splines.

These limitations are not as narrow as one might think. Straight-line trajectories are easily executed by robot arms, and especially cartesian arms. Thus the one-dimensional problem could represent a fragment of a spatial trajectory for the robot hand, and along this trajectory the 'pursuit functions' could represent interference by other robot arms. On the other hand, the spacetime curve represented by these interfering arms is unlikely to be a quadratic spline. The curve is probably better represented by piecewise linear trajectories with infinite acceleration where the velocity changes. However the methods in this paper can obviously be extended easily to handle this situation. For cartesian arms, any other trajectory is artificial, and this is convenient for us, since it is clear from this paper that we can handle

the interaction of linear and quadratic spacetime curves without encountering polynomials of unpleasantly high degree.

A more realistic direction of research would be to consider the problem as being defined along the boundary of a circle rather than a line-segment, since such motions are quite common in commercial arms. The problem of handling centrifugal force leads to a new domain with interesting possibilities.

The motion-planning problem considered here was a feasibility problem. What about the optimization problem, i.e., given the requirements (2.0, 2.1), what is the minimal M such that (2.2) can also be satisfied? Actually, the methods of this paper are suitable for attacking such a problem. Suppose that p represents the initial position and velocity. We know the structure of $R_{a,b}(p)$. Suppose that we regard this set as being parametrized by M. Clearly, it increases (in terms of set-inclusion) as M increases, and its structure remains qualitatively unchanged except at a bounded number of 'critical values' of M. It is also clear that supposing q to be the required final position and velocity, M is optimal if and only if q lies on the boundary of $R_{a,b}(p)$. Thus, if we knew qualitatively how $R_{a,b}(p)$ should look like at the optimal value of M, it should be quite easy to compute the optimal value.

Certainly this program can be carried out for the case of a single pursuit function g (as if $f = -\infty$). Then for any M the lower boundary of $R_{a,b}(p)$, supposing it to be nonempty, is a parabola with second derivative $(2M)^{-1}$, and the upper boundary can be computed by the methods of Appendix B. Clearly, also, the criticalities are those values of M for which there exists a trajectory with acceleration $-M$ meeting g at more than two points, and there are clearly at most n^3 such criticalities. (This is a very loose upper bound.) For this case, therefore, it would not be difficult to generate a polynomial-time solution to the optimization problem. The more general case seems rather more difficult, since the process of 'shading,' when repeated sufficiently often, can result in quite complex chains of dependencies between the two pursuit functions. It is not clear what difficulties this might raise, nor even whether these dependencies could increase the algebraic complexity of the problem (the most complex operation we have needed in this paper is the exact solution of quadratic equations).

Acknowledgements. It is a pleasure to thank Robert Burridge, Alan Siegel, and Chee-Keng Yap for stimulating discussions of this problem.

9. Appendix A: calculation of bang-bang interpolant

In this appendix we compute exactly, or at least give the algebraic equations governing, the two quantities τ and μ which characterize the unique minimax acceleration path from

(a, x_a, v_a) to (b, x_b, v_b), i.e., the 'bang-bang interpolant.' Here τ is the time where the acceleration changes, v_τ is the velocity at this time, and μ is the initial acceleration. These quantities are related by the following equations. First, the accelerations balance:

$$\frac{v_\tau - v_a}{\tau - a} = \frac{v_b - v_\tau}{\tau - b}$$

i.e.

$$\frac{\tau - a}{v_\tau - v_a} + \frac{\tau - b}{v_\tau - v_b} = 0; \tag{A1}$$

Second, the trajectory passes through x_b at time b:

$$\frac{(\tau - a)(v_\tau + v_a)}{2} + \frac{(b - \tau)(v_\tau + v_b)}{2} = x_b - x_a,$$

i.e.,

$$(\tau - a)(v_\tau + v_a) - (\tau - b)(v_\tau + v_b) = 2(x_b - x_a). \tag{A2}$$

This can be written as

$$\begin{bmatrix} (v_\tau + v_a) & -(v_\tau + v_b) \\ \dfrac{1}{v_\tau - v_a} & \dfrac{1}{v_\tau - v_b} \end{bmatrix} \begin{bmatrix} \tau - a \\ \tau - b \end{bmatrix} = \begin{bmatrix} 2(x_b - x_a) \\ 0 \end{bmatrix}. \tag{A3}$$

The determinant of this system of equations is

$$\Delta = \frac{2v_\tau^2 - v_a^2 - v_b^2}{(v_\tau - v_a)(v_\tau - v_b)} \tag{A4}$$

First, let us assume nondegeneracy, i.e., that Δ be nonzero. Cramer's rule implies

$$(\tau - a)\Delta = \frac{2(x_b - x_a)}{v_\tau - v_b} \tag{A5}$$

and

$$(\tau - b)\Delta = \frac{-2(x_b - x_a)}{v_\tau - v_a}. \tag{A6}$$

Subtracting, substituting for Δ, and multiplying through by $(v_\tau - v_b)(v_\tau - v_a)$, we obtain

$$(b - a)(2v_\tau^2 - v_a^2 - v_b^2) = 2(x_b - x_a)(2v_\tau - v_a - v_b) \tag{A7}$$

or

$$v_\tau^2 - \frac{v_a^2+v_b^2}{2} = C(2v_\tau - v_a - v_b), \tag{A8}$$

where

$$C = \frac{x_b - x_a}{b-a};$$

and this is easily simplified (completing the square) to

$$(v_\tau - C)^2 = \tfrac{1}{2}(v_a - C)^2 + \tfrac{1}{2}(v_b - C)^2. \tag{A9}$$

Second, if one examines the 'degenerate' case, where $x_a = x_b$ and consequently the vector to the right of equation A3 is zero, existence of a nontrivial solution implies that the determinant Δ must be zero, and that condition is easily seen to lead again to equation A9 (in this case, C is zero).

Before continuing along these lines, let us consider what happens when a linear or parabolic solution can be fitted to the integration constraints, so we should expect both v_τ and τ to be ill-defined. (It is clear from Lemma 1 that this is the optimal solution.) This occurs precisely when

$$\frac{x_b - x_a}{b-a} = \tfrac{1}{2}(v_b + v_a), \tag{A10}$$

in which case the constant C in equation (A9) can be rewrittten, and the equation becomes, after simplification,

$$\left(v_\tau - \frac{v_a+v_b}{2}\right)^2 = \left(\frac{v_b - v_a}{2}\right)^2$$

with the two solutions $v_\tau = v_a$ or v_b; which is to be expected. Next let us consider the case where equation (A9) has the solution $v_\tau = v_a$: this is possible when (i) $v_b = v_a$ or (ii) $v_b = 2C - v_a$. Note that case (ii) is merely a restatement of equation (A10) and has already been covered: the other solution is $v_\tau = v_b$. In case (i), equation (A1) has the solution $\tau = \tfrac{1}{2}(a+b)$, which clearly yields the optimal interpolant.

Let us return to equation A5 which defines $\tau - a$ in terms of v_τ. This simplifies to

$$(\tau - a)(2v_\tau^2 - v_a^2 - v_b^2) = 2(x_b - x_a)(v_\tau - v_a).$$

If we use equation A7 we can deduce that

$$\frac{(\tau-a)(2v_\tau-v_a-v_b)}{(b-a)} = v_\tau-v_a$$

and therefore

$$\beta = \frac{v_\tau-v_a}{2v_\tau-v_a-v_b}, \tag{A11}$$

where

$$\beta = \frac{\tau-a}{b-a}.$$

First let us consider the situation where equation (A11) is ill-defined, so $v_\tau = \pm \frac{1}{2}(v_a+v_b)$. Then equation (A8) implies

$$(v_a+v_b)^2-2(v_a^2+v_b^2) = 0,$$

so $v_a = v_b$ again, and this equals v_τ by assumption. Equation (A1) gives no information at all but equation (A2) becomes

$$v_a(b-a)=x_b-x_a,$$

which means that the bang-bang interpolant has uniform velocity and the acceleration is zero.

Equation A9 is quadratic and therefore has two solutions in general; however, we are only interested in those solutions for which β is between 0 and 1. If one revises equation A11 in terms of equation A9, replacing v_τ by $v_\tau-C$ and so forth, it turns out that

$$\beta = \frac{1}{1+\frac{w-v}{w-u}} \tag{A12}$$

where $u = v_a-C$, $v = v_b-C$, and $w = v_\tau-C = \pm\sqrt{\frac{1}{2}(u^2+v^2)}$ according to equation A9. The previous discussion has covered the case where equation (A12) does not have a well-defined solution. This corresponds to the situation where $|u| = |v|$. For the other cases, it is easy to see that exactly one of the two solutions to equation (A9) defines a meaningful value for τ – i.e., one for which β lies between 0 and 1, as follows. For equation (A12) to define a legitimate value for β it is necessary and sufficient that $w-v$ and $w-u$ have the same sign. Supposing that $|u| \neq |v|$, and observing that the criterion is symmetric in u and v, we may suppose for simplicity that $|v| > |u|$. Consequently, $|u| < |w| < |v|$. If v is positive then we must choose the negative value for w, since the positive value would force $w-u$ to be positive and $w-v$ negative. But if v is negative we must choose the positive value for w. The case $|u| = |v|$, which has been discussed above, is subject to a similar easy analysis. Summarizing this analysis, we have the following lemma (Lemma 2 restated).

Except when equation (A2) admits a uniform-acceleration solution, there is exactly one solution v_τ to equation (A9) which defines an admissible solution to equation (A1), and the rule defining it is:

$$v_\tau = C - s\sqrt{\frac{(v_a - C)^2 + (v_b - C)^2}{2}},$$

where

$$C = \frac{x_b - x_a}{b - a},$$

and $s = +1$ if that one of $v_a - C$, $v_b - C$ which has larger absolute value is positive, else $s = -1$. (When both of these quantities have the same absolute value but opposite sign, (A2) admits a uniform-acceleration solution.)

10. Appendix B: nature of upper boundary of phase region formed by g.

This appendix studies the curve $v_\ell(x)$ defined in equation (4.2). This is defined so that the trajectory $\delta_{(b,x,v)}$ touches g (or the associated set G) for $v = v_\ell(x)$, at some point τ, say, but does not cross G.

It is easiest to work with the *latest* point τ where the indicated trajectory meets G. Clearly, this point is nondecreasing as x increases (this can be shown by a perturbation argument). Also, except for finitely many values of x, there is exactly one such point of tangency. Thus the following formulation is satisfactory.

Suppose that $\ddot{g} > -M$ at τ and τ is not a point where it changes. We consider the points

$$x(\tau) = \delta_{(\tau, g(\tau), \dot{g}(\tau))}(b)$$

and

$$v(\tau) = \dot{\delta}_{(\tau, g(\tau), \dot{g}(\tau))}(b)$$

Let us write

$$\delta_{(\tau, g(\tau), \dot{g}(\tau))}(t) = A - ½M(t - B)^2$$

Clearly,

$$g(\tau) = A - ½M(\tau - B)^2,$$

and

$$\dot{g}(\tau) = M(B-\tau)$$

so

$$B = \tau + \frac{\dot{g}(\tau)}{M}$$

and

$$A = g(\tau) + \frac{1}{2M}\dot{g}(\tau)^2.$$

We can substitute these to get an explicit formulation of $(x(\tau), v(\tau))$.

$$(x(\tau), v(\tau)) = \left(g(\tau) + \frac{1}{2M}\dot{g}(\tau)^2 - \tfrac{1}{2}M(b-\tau-\frac{1}{M}\dot{g}(\tau))^2, M(\tau+\frac{1}{M}\dot{g}(\tau)-b)\right).$$

Differentiating with respect to τ we obtain

$$\left(\dot{g}(\tau) + \frac{1}{M}\dot{g}(\tau)\ddot{g}(\tau) + M(b-\tau-\frac{1}{M}\dot{g}(\tau))(1+\frac{1}{M}\ddot{g}(\tau)), M(1+\frac{1}{M}\ddot{g}(\tau))\right)$$

which simplifies to

$$\left(M(b-\tau)(1+\frac{1}{M}\ddot{g}(\tau)), M(1+\frac{1}{M}\ddot{g}(\tau))\right).$$

In other words, since $M + \ddot{g}(\tau) > 0$,

$$\frac{d}{d\tau}v(\tau) > 0$$

and

$$\frac{d}{d\tau}x(\tau) = (b-\tau)\frac{dv}{d\tau}.$$

Since v is strictly increasing with τ we can regard it as the independent variable in which case we have

$$\frac{dx}{dv} = (b-\tau) > 0$$

Note that dx/dv decreases as τ increases. This means that as B moves from left to right, and τ changes continuously, dx/dv decreases continuously, and at discontinuous jumps of τ dx/dv is discontinuous but decreases. To compute d^2x/dv^2 explicitly, we write

$$\frac{d^2x}{dv^2} = \frac{d}{dv}\left(\frac{dx}{dv}\right) = \frac{d}{dv}\left(\frac{dx}{d\tau}\frac{d\tau}{dv}\right) = \frac{d}{dv}(b-\tau) = -\frac{d\tau}{dv}.$$

Or, explicitly,

$$\frac{d^2x}{dv^2} = \frac{-1}{M+\ddot{g}(\tau)}.$$

which is constant and negative if $\ddot{g}$ is constant and $\geq -M$. This implies that the curve traced out is monotonic in v, piecewise quadratic, with negative second derivative, and only downward jumps in the first derivative. Note that as τ converges to b, $x(\tau)$ converges to $g(b)$ (assuming that $\ddot{g} > -M$ at b) and dx/dv converges to zero.

As a last point one should note that when the point τ of tangency is at a 'convex corner' of g, supposing such to be defined for g (for instance, at the endpoints a and b), it is clear that τ remains constant for some time (as B increases) and therefore, following much the same analysis with some care, dx/dv remains constant.

11. Appendix C: growth properties of $X(\tau)$.

Here we want to analyze the way the set

$$X(\tau) = FR_{a,\tau}{}^{-1}(\tau, g(\tau), \dot{g}(\tau)) \tag{B1}$$

changes as τ moves rightwards from a. The function g is as follows:

$$g(\tau) = G + \tfrac{1}{2}N(\tau - H)^2, \tag{B2}$$

where G and H are fixed parameters and $N > -M$. First let $S(\tau)$ represent $FR_{0,\tau}{}^{-1}(0,0)$:

$$S(\tau) = \{(x,v)\colon (v-M\tau)^2 - 2M^2\tau^2 \leq 4Mx \leq 2M^2\tau^2 - (v+M\tau)^2\}. \tag{B3}$$

Then in the notation of Minkowski sums, if $(x(\tau), v(\tau))$ represents the point lying, along the straight-line trajectory which touches g at time τ, at time 0,

$$X(\tau) = (x(\tau), v(\tau)) + S(\tau). \tag{B4}$$

Now the straight-line trajectory touching g at τ is, explicitly,

$$g(\tau) + (t-\tau)\dot{g}(\tau).$$

By a simple calculation we deduce that

$$x(\tau) = G + \tfrac{1}{2}NH^2 - \tfrac{1}{2}N\tau^2 \quad \text{and} \quad v(\tau) = N(\tau - H).$$

Thus equation (B4) becomes

$$X(\tau) = (G + \tfrac{1}{2}NH^2, -NH) + (-\tfrac{1}{2}N\tau^2, N\tau) + S(\tau). \tag{B5}$$

We consider the whole parabola containing its lower boundary: this is the point-set

$$(G + \tfrac{1}{2}NH^2, -NH) + (-\tfrac{1}{2}N\tau^2, N\tau) + (-\tfrac{1}{2}M\tau^2, M\tau) + \{(x,v): x = \frac{v^2}{4M}\}. \tag{B6}$$

Let us differentiate, with respect to τ, the curve traced out by the apex B of this moving parabola:

$$\frac{dB}{d\tau} = (M+N)(-\tau, 1).$$

To partition the moving parabola into an 'advancing' and 'receding' part, we locate that point on the parabola where the tangent is parallel to this direction of motion. The parabola can be parametrized in the form

$$\left(\frac{v^2}{4M}, v\right) \tag{B7}$$

which we can differentiate with respect to v, getting the vector direction

$$\left(\frac{v}{2M}, 1\right)$$

which is parallel to $dB/d\tau$ when $v = -2M\tau$. Therefore, the 'stagnation' point on the curve is that at displacement

$$(M\tau^2, -2M\tau)$$

from the moving apex, which is the upper corner A of $X(\tau)$.

If $N > M$ then the direction of motion of this upper corner is in the same sense as $(-\tau, 1)$, i.e., south-eastwards; also, the lower corner B is moving southeast; therefore the set Y_2 has the indicated growth properties.

If $N = M$ then it is clear that the upper boundary of the sets $X(\tau)$ is confined to the same parabolic curve. If $N < M$ then the direction of motion of the upper corner is in the same sense as $(\tau, -1)$, i.e., northwest, and the set Y_1 has the indicated growth properties in this case.

This covers the properties set out in Lemma 12.

12. REFERENCES

1. R. Farwig and D. Zwick, "A variation on the Landau Problem," *American Mathematical Monthly* **93:6** pp. 444-451 (June-July 1986).

NYU COMPSCI TR-230
O'Dunlaing, Colm
Motion-planning with
inertial constraints.
c.2